



Extremal Problems on the Hypercube

Trevor Alvaro Anthony Pinto

Submitted in partial fulfilment of the requirements of the degree
of Doctor of Philosophy

March 2016

Statement of Originality

I, Trevor Alvaro Anthony Pinto, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Trevor Alvaro Anthony Pinto.

Date: 22 March 2016

Details of collaboration and publications:

Chapter 2 is based on work with my supervisor, J. Robert Johnson, the majority of which we wrote the following preprint [49], which has now been accepted for publication in Combinatorics, Probability and Computing. We worked closely together on the material of this chapter, all of it being joint work.

The contents of Chapter 3 are largely based on the preprint [82], but has also some

additional material, namely the work on grids in Section 3.5.

The fourth chapter is based almost exclusively on published work [81].

The final chapter is the result of a collaboration with Klas Markström of Umeå University.

We worked closely together on the material of this chapter, all of it being joint work.

Abstract

The hypercube, Q_d , is a natural and much studied combinatorial object, and we discuss various extremal problems related to it.

A subgraph of the hypercube is said to be (Q_d, F) -saturated if it contains no copies of F , but adding any edge forms a copy of F . We write $\text{sat}(Q_d, F)$ for the saturation number, that is, the least number of edges a (Q_d, F) -saturated graph may have. We prove the upper bound $\text{sat}(Q_d, Q_2) < 10 \cdot 2^d$, which strongly disproves a conjecture of Santolupo that $\text{sat}(Q_d, Q_2) = (\frac{1}{4} + o(1)) d 2^{d-1}$. We also prove upper bounds on $\text{sat}(Q_d, Q_m)$ for general m .

Given a down-set A and an up-set B in the hypercube, Bollobás and Leader conjectured a lower bound on the number of edge-disjoint paths between A and B in the directed hypercube. Using an unusual form of the compression argument, we confirm the conjecture by reducing the problem to the case of the undirected hypercube. We also prove an analogous conjecture for vertex-disjoint paths using the same techniques, and extend both results to the grid.

Additionally, we deal with subcube intersection graphs, answering a question of Johnson and Markström of the least $r = r(n)$ for which all graphs on n vertices may be represented as subcube intersection graph where each subcube has dimension exactly r . We also contribute to the related area of biclique covers and partitions, and study relationships between various parameters linked to such covers and partitions.

Finally, we study topological properties of uniformly random simplicial complexes, employing a characterisation due to Korshunov of almost all down-sets in the hypercube as a key tool.

Contents

1	Introduction	10
1.1	The Hypercube	10
1.2	Saturated subgraphs of the hypercube	12
1.3	Vertex and Edge-Disjoint Paths in the Cube	13
1.4	Biclique Covers and Partitions	13
1.5	Uniformly Random Simplicial Complexes	14
2	Saturated Subgraphs of the Hypercube Graph	16
2.1	Introduction	17
2.2	Preliminaries	22
2.3	Zero Density Bound on $\text{sat}(Q_d, Q_m)$	22
2.4	Saturated Hypergraphs	28
2.5	Bounded Average Degree Constructions	30
2.5.1	Semi-saturation	30
2.5.2	Improved bound for $\text{sat}(Q_d, Q_2)$	31
2.6	Lower Bounds	35
2.7	The (Q_d, Q_2) -free process	36
2.8	Discussion	39
2.8.1	(Q_d, F) -saturation	39
2.8.2	Semi-saturation and weak-saturation	40
2.8.3	Vertex saturation	41

3	Directed Paths in the Cube	42
3.1	Introduction	43
3.1.1	Edge-Disjoint Paths	43
3.1.2	Vertex-Disjoint Paths	46
3.2	Flow Theorems	48
3.3	Directed Isoperimetric Inequalities and Directed Paths	51
3.3.1	Compression Operators	51
3.3.2	Edge Version of Main Result	52
3.3.3	Vertex Version of Main Result	54
3.4	Bollobás and Leader’s Theorems	55
3.4.1	Undirected Edge-Disjoint Paths	55
3.4.2	Matchings and Undirected Vertex-Disjoint Paths	56
3.5	Generalisation to Grids	60
3.5.1	Edge-Disjoint Paths in a Grid	61
3.5.2	Vertex-Disjoint Paths in a Grid	63
4	Local Biclique Decompositions	64
4.1	Introduction	65
4.1.1	Biclique covers and partitions	65
4.1.2	Subcube Intersection Graph Representations	67
4.2	Relationship between τ and ρ	68
4.3	Bounding bp in terms of bc	69
4.4	Relationship between lbc and lbp	72
4.5	Bounding $\text{lbp}(G)$, for $\text{lbc}(G) = 2$	75
4.6	Discussion	77
4.6.1	Complexity of Boolean Matrices	78
5	Uniformly Random Simplicial Complexes	81
5.1	Introduction	82
5.1.1	Korshunov’s characterization of simplicial complexes	86
5.1.2	Summary of models	88

5.2	Preliminaries	89
5.2.1	The Probabilistic Method	89
5.2.2	Asymptotic Results	90
5.3	Complete Skeletons	91
5.4	Homology Group	94
5.5	Euler Characteristic	98
5.6	Shellability and the h -Vector	100
5.7	Evasiveness	104

Acknowledgements

First and foremost, I would like to thank my supervisor J. Robert Johnson for the invaluable support, both mathematical and moral, he has given me throughout my PhD. His patience, diligence and encouragement helped me enormously during both the research and the writing stages of my PhD and without him my PhD would have been far less enjoyable and fruitful.

I must also thank Mark Walters and David Ellis for the many different mathematical insights and useful pieces of advice they provided me with. More generally, I would like to thank the many mathematicians at Queen Mary University of London and outside, in particular Simon Griffiths, for useful discussions, thoughts and suggestions.

I would also like to thank the current and former staff, academic and support, of Queen Mary, University of London, in particular the members and organisers of the Combinatorics Study Group, for helping to provide an environment that was both pleasant and stimulating for research.

Klas Markström deserves special thanks for hosting me on a very fruitful and interesting visit of Umeå University in Sweden, as do the various other academic staff and students who gave me a very warm welcome. I would also like to thank Queen Mary's Eileen Colyer Prize, and the two anonymous members of staff who very generously funded it from their own pockets, for the financial assistance that made this visit possible.

I am also grateful for the EPSRC doctoral studentship that allowed me to undertake my PhD.

On a more personal level, I am grateful to my parents and my siblings for the incredible amounts of support and understanding (and grammar advice!) they gave me, especially

during the writing stage of my PhD.

I am also thankful to all my friends among the Queen Mary PhD students for providing a real sense of community that made my time at Queen Mary a very enjoyable experience.

Last but not least, I would like to thank the many other friends, colleagues and family members who have given me advice, support and encouragement during all the bumps in the road of my PhD journey, as well as for all the good times we have shared.

Chapter 1

Introduction

1.1 The Hypercube

The d -dimensional hypercube, Q_d , is an object of fundamental importance in graph theory and combinatorics. One of the reasons for its importance is the large number of natural viewpoints in which it arises. We will primarily consider it as a graph with vertex set $\{0, 1\}^d$ and with edges between each pair of vertices that differ in exactly one coordinate. It can be convenient to treat the vertex set as \mathbb{F}_2^d , the d -dimensional vector space over the field with 2 elements. We write e_1, \dots, e_d for the canonical basis of \mathbb{F}_2^d (e_i is the vector with a 1 in the i^{th} coordinate, and 0's elsewhere). We can see that x is adjacent to y if and only if $y = x + e_i$, for some $i \in \{1, \dots, d\}$.

Alternatively, we may view the vertex set of Q_d as $\mathcal{P}[d]$, the power set of the d element set. In this viewpoint, the set x is adjacent to the set y if and only if $|x \triangle y| = 1$. This formulation highlights the utility of the hypercube as a natural way of studying set systems. When using this notation, we will often shorten $\{i\}$ to i and, more generally, $\{i_1, i_2, \dots, i_t\}$ to $i_1 i_2 \dots i_t$. Since every set is Q_d -adjacent to its maximal proper subsets, Q_d is the Hasse diagram for the poset of $\mathcal{P}[d]$ under the inclusion relation.

A *subcube* of Q_d is an induced subgraph isomorphic to Q_m , for some $m \leq d$. It is easy to see that a set of vertices, S , is the vertex set of a subcube if and only if there is some set of coordinates $J \subseteq [d] = \{1, 2, 3, \dots, d\}$, and constants $a_j \in \{0, 1\}$ for each $j \in J$ such that

$(x_1, \dots, x_d) \in S$ if and only if for all $j \in J$, $x_j = a_j$. *Fixed coordinates* are those coordinates in J , whereas *free coordinates* are coordinates that are not fixed. We can thus represent a subcube as an element of $\{0, 1, *\}^d$, with stars in the free coordinates, and a_j in the fixed coordinates. As edges can be thought of as Q_1 's, we may also represent edges as elements of $\{0, 1, *\}^d$ in this way. We will say an edge or subcube lies *along* the directions i_1, \dots, i_k if these contain all the free coordinates of the edge or subcube.

A simple use of this notation is for counting the number of edges, or more generally the number of dimension m subcubes in Q_d . Indeed, there are $\binom{d}{m}$ ways of choosing where to place the m stars and 2^{d-m} ways of choosing either 0 or 1 in the other coordinates. Thus there are $\binom{d}{m} 2^{d-m}$ dimension m subcubes in Q_d ; in particular, Q_d has $d2^{d-1}$ edges.

The *weight* of $x \in V(Q_d)$ is the number of coordinates of x that are 1. It is often useful to partition the vertices of Q_d into layers, where the i^{th} layer consists of vertices with weight i and edges lie only between adjacent layers. In set notation, the i^{th} layer is $[d]^{(i)}$, that is, the set of all subsets of $[d]$ of size i .

We may write $Q_{d_1+d_2}$ as $Q_{d_1} \square Q_{d_2}$, the graph Cartesian product of Q_{d_1} and Q_{d_2} . In other words, $Q_{d_1+d_2}$ is formed by replacing each vertex of Q_{d_2} with a copy of Q_{d_1} . We call these copies *principal Q_{d_1} 's*. Where there was a Q_{d_2} edge, e , we instead put edges between corresponding vertices of the principal Q_{d_1} 's placed at the endpoints of e . So we have two types of edges: *internal edges* which have both endpoints in the same principal Q_{d_1} and *external edges* which have endpoints in different principal Q_{d_1} 's. Notice that there are d_1 directions along which internal edges lie, and d_2 directions along which external edges lie. This view of $Q_{d_1+d_2}$ is useful for induction and we will make heavy use of it in Chapter 2; we will write $Q_{d_1+d_2}$ as $Q_{d_1} \square Q_{d_2}$ when we wish to use this viewpoint.

A natural generalisation of the hypercube is the *grid* graph, P_m^d . It is the graph with vertex set $[m]^d$ and an edge between two vertices (x_1, x_2, \dots, x_d) and (y_1, y_2, \dots, y_d) if and only if there is some i such that $|x_i - y_i| = 1$, and $x_j = y_j$ for all $j \neq i$. Note that P_2^d is isomorphic to Q_d and that P_m^d is the Cartesian product of d copies of the path with m vertices.

1.2 Saturated subgraphs of the hypercube

Given graphs H and F , we say a graph is (H, F) -saturated if it is a maximal F -free subgraph of H . A graph is said to be (H, F) -semi-saturated if it is a subgraph of H and adding any edge forms a new copy of F . Saturated and semi-saturated graphs have been studied since Zykov [93] in the 1940s, and again by many other authors. We are particularly interested in the case where $H = Q_d$ and $F = Q_m$. The minimum number of edges a (Q_d, Q_m) -saturated graph (resp. (Q_d, Q_m) -semi-saturated graph) can have is denoted by $\text{sat}(Q_d, Q_m)$ (resp. $s\text{-sat}(Q_d, Q_m)$). We prove that $\lim_{d \rightarrow \infty} \frac{\text{sat}(Q_d, Q_m)}{e(Q_d)} = 0$, for fixed m , strongly disproving a conjecture of Santolupo (reported in a survey by J. Faudree, R. Faudree and Schmitt [35]) that, when $m = 2$, this limit is $\frac{1}{4}$. Additionally, we show that our approach may be used in the more general setting, in which we consider the hypercube as a regular hypergraph.

Using a different method, we show that $s\text{-sat}(Q_d, Q_m) = O(2^d)$, for fixed m . In the case $m = 2$, we are able to amend the proof to get an analogous result for saturation, i.e. that $\text{sat}(Q_d, Q_2) = O(2^d)$. We also prove the lower bound $s\text{-sat}(Q_d, Q_m) \geq \frac{m+1}{2} \cdot 2^d$, thus determining the order of $\text{sat}(Q_d, Q_2)$.

Saturated graphs have a natural connection to constrained random processes. In particular, we look at the (Q_d, Q_2) -free process. This is a random process consisting of subgraphs G_0, \dots, G_M of Q_d such that G_0 is empty and that G_{i+1} is formed by adding to G_i an edge chosen uniformly at random subject to the condition that no copy of Q_2 is formed when it is added. The process stops when no further edge may be added, that is, when G_M is (Q_d, Q_2) -saturated, and so it may be considered a random greedy algorithm for constructing a (Q_d, Q_2) -saturated graph. This process is analogous to the well-known triangle-free process, see for instance [11] or [83]. We show that with high probability, the number of edges in the saturated graph produced is at least $cd^{2/3}2^d$, for some constant c . We also discuss reasons why the methods applied to the triangle-free process by [11] and [83] seem difficult to use for this problem.

1.3 Vertex and Edge-Disjoint Paths in the Cube

Let A and B be disjoint sets of vertices of Q_d . We write $p_e(Q_d, A, B)$ for the size of the largest family of edge-disjoint paths between A and B . In 1997, Bollobás and Leader proved that if $|A| = |B| = 2^k$, then $p_e(Q_d, A, B) \geq (d - k)2^k$. This result can be seen to be tight by considering k -dimensional subcubes. We write $p_e(\vec{Q}_d, A, B)$ for the size of the largest family of edge-disjoint directed paths between A and B , that is, paths whose vertices form a chain. Bollobás and Leader conjectured that when A is a down-set, B is an up-set and $|A| = |B| = 2^k$, the same bound holds for directed paths, that is $p_e(\vec{Q}_d, A, B) \geq (d - k)2^k$.

We use a novel type of compression argument to prove a stronger version of this conjecture. More precisely, for any down-set A and up-set B of arbitrary cardinality, the size of the largest family of edge-disjoint paths between a down-set A and an up-set B is the same as the size of the largest family of directed edge-disjoint paths between A and B . In other words, $p_e(\vec{Q}_d, A, B) = p_e(Q_d, A, B)$.

Bollobás and Leader made an analogous conjecture for paths with vertex-disjoint interiors and we prove a strengthening of this by similar methods. We also generalize both results to the setting where A and B are sets of vertices in the grid P_m^d .

1.4 Biclique Covers and Partitions

The *biclique cover number* (resp. *biclique partition number*) of a graph G , $\text{bc}(G)$ (resp. $\text{bp}(G)$), is the least number of bicliques—complete bipartite subgraphs—that are needed to cover (resp. partition) the edges of G .

The *local biclique cover number* (resp. *local biclique partition number*) of a graph G , $\text{lbc}(G)$ (resp. $\text{lbp}(G)$), is the least r such that there is a cover (resp. partition) of the edges of G by bicliques with no vertex in more than r of these bicliques.

We show that $\text{bp}(G)$ may be bounded in terms of $\text{bc}(G)$, in particular, $\text{bp}(G) \leq \frac{1}{2}(3^{\text{bc}(G)} - 1)$. However, the analogous result does not hold for the local measures. Indeed, in our main result, we show that $\text{lbp}(G)$ can be arbitrarily large, even for graphs with $\text{lbc}(G) = 2$. For such graphs, G , we try to bound $\text{lbp}(G)$ in terms of additional information about biclique covers of G . We both answer and leave open questions related to this.

There is a well-known link between biclique covers and subcube intersection graphs. We consider the problem of finding the least $r(n)$ for which every graph on n vertices can be represented as a subcube intersection graph in which every subcube has dimension r . We reduce this problem to the much studied question of finding the least $d(n)$ such that every graph on n vertices is the intersection graph of subcubes of a d -dimensional cube.

1.5 Uniformly Random Simplicial Complexes

Random structures have long played an important role both in combinatorics and in wider mathematics, in part because they often provide good bounds for extremal questions and also because they give an insight into the typical behaviour of mathematical objects. Much-studied random structures include random graphs (see for instance [15]) and random matrices (see for instance [74]).

In recent years, random simplicial complexes have received much attention, in particular the model introduced by Linial and Meshulam in [66]. For a fixed constant, t , they generate simplicial complexes on n vertices by starting with a complete $(t - 1)$ -skeleton and adding t -sets uniformly independently at random, with probability p . Topological properties of this model have been studied further by many authors, for instance see [73], [4], [6] and [40].

Despite being very natural, the uniformly random simplicial complex, $\mathcal{U}(n)$ has received rather less attention. We say $\Delta \sim \mathcal{U}(n)$ if it is chosen uniformly from the collection of all simplicial complexes on n vertices (here we view simplicial complexes on n vertices purely as collections of subsets of $[n]$ (and so homeomorphic complexes are considered to be distinct). We show how a combinatorial classification of almost all simplicial complexes due to Korshunov [59] may be used to prove topological properties of $\mathcal{U}(n)$. Among other results, we show that if $\Delta \sim \mathcal{U}(n)$ then with high probability its top homology group is non-trivial and also bound its Euler Characteristic.

Another random model that we propose is the random pure model $\mathcal{RP}(n, t, p)$, in which sets of size t are chosen independently with probability p to be maximal faces of the simplicial complex. We again study topological properties of this model, both when t is a constant and when it is a fixed proportion of n , and note similarities between it and the uniformly

random simplicial complex.

We also note our original motivation for studying $\mathcal{U}(n)$, a property of Boolean functions known as evasiveness. Our work on the uniform random complex has a very easy corollary in this area. Although there may be a more direct way of deriving the result, we feel it shows the utility of the uniform random complex.

Chapter 2

Saturated Subgraphs of the Hypercube Graph

2.1 Introduction

Let F be a graph. We say that a graph, G , is F -free if it contains no subgraphs isomorphic to F . If G is a maximal F -free subgraph of H , we say that G is (H, F) -saturated. In other words, G is (H, F) -saturated if it is a subgraph of H , it is F -free and the addition of any edge from $E(H) \setminus E(G)$ forms a copy of F . In this context, H is referred to as the *host graph*, F as the *forbidden graph* and G as a *saturated graph*.

The famous Turán problem in extremal combinatorics can be expressed naturally in the language of saturated graphs. The extremal number of F , written $ex(K_n, F)$ and often shortened to $ex(n, F)$, is usually defined as the maximum number of edges in an F -free subgraph of K_n . However, it can equivalently be written as:

$$ex(K_n, F) = \max\{e(G) : G \text{ is } (K_n, F)\text{-saturated}\}.$$

This formulation yields a natural opposite to the Turán problem. We define the *saturation number of F* , $sat(H, F)$, as:

$$sat(H, F) = \min\{e(G) : G \text{ is } (H, F)\text{-saturated}\}.$$

A variant of this is the *semi-saturation number*, written $s\text{-}sat(H, F)$. We say that a graph, G , is (H, F) -semi-saturated if G is a subgraph of H and adding any edge from $E(H) \setminus E(G)$ increases the number of copies of F . Thus a graph is (H, F) -saturated if and only if it is (H, F) -semi-saturated and F -free. We define:

$$s\text{-}sat(H, F) = \min\{e(G) : G \text{ is } (H, F)\text{-semi-saturated}\}.$$

Perhaps surprisingly, $s\text{-}sat(H, F)$ may be significantly smaller than $sat(H, F)$. In other words, to ensure that adding an edge always produces a copy of F , it may be optimal to start with a graph that contains a copy of F . Indeed, this occurs even for the relatively simple case of $H = K_n$ and $F = P_k$, the path with k edges, see [56].

The most frequently studied host graph is the complete graph, K_n . Ever since work in the area began with Zykov [93] and Erdős, Hajnal and Moon [33], the numbers $s\text{-}sat(K_n, F)$ and $sat(K_n, F)$ have received much attention: see for instance the survey articles of Pikhurko [80] and of J. Faudree, R. Faudree and Schmitt [35] and the references contained therein.

Other host graphs that have been studied include complete bipartite graphs, see for instance the works of Wessel [91] and Bollobás [14], and Erdős-Renyi random graphs, as in recent work by Korándi and Sudakov [62].

In the literature, $\text{sat}(K_n, F)$ is often written as $\text{sat}(n, F)$ and (K_n, F) -saturated is usually written as F -saturated. Since the results in this chapter concern a different host graph, we will reserve this latter abbreviation for a different meaning.

A much studied variant of the Turán problem was initiated by Erdős in [31] and expanded upon by Alon, Krech and Szabò [2]. For a fixed graph F , they ask for $\text{ex}(Q_d, F)$, the maximum number of edges in an F -free subgraph of the d -dimensional hypercube, Q_d . The most natural case is $F = Q_m$, a fixed dimension hypercube. This is wide open, even for the case $m = 2$. The asymptotic edge density of a maximum Q_2 -free graph, i.e. $\lim_{d \rightarrow \infty} \frac{\text{ex}(Q_d, Q_2)}{e(Q_d)}$ was conjectured by Erdős [31] to be $\frac{1}{2}$, with \$100 offered for a solution. In other words, the graph whose edges are all those with lowest weight endpoint having even weight is conjectured to be, asymptotically, the best possible. The conjecture is still open, despite the attention of many authors—see for instance the work of Balogh, Hu, Lidický and Liu [8] and of Brass, Harborth and Nienborg [18].

In this chapter, we focus mainly on the saturation and semi-saturation problems where the host graph is the hypercube and the forbidden graph is a subcube. That is, we study $\text{sat}(Q_d, Q_m)$ and $s\text{-sat}(Q_d, Q_m)$. For brevity, we shall often write F -saturated (resp. F -semi-saturated) rather than (Q_d, F) -saturated (resp. (Q_d, F) -semi-saturated) in the remainder of this chapter, when the value of d is clear or irrelevant.

The first result of this sort, due to Nieves Roman [79], is a construction of a family of (Q_n, Q_2) -saturated graphs with asymptotic edge density $\frac{5}{16}$, in the limit $n \rightarrow \infty$. The best result along these lines is that of Choi and Guan [23], which improves $\frac{5}{16}$ to $\frac{1}{4}$. That is, they show:

$$\lim_{d \rightarrow \infty} \frac{\text{sat}(Q_d, Q_2)}{e(Q_d)} \leq \frac{1}{4}.$$

A conjecture that this is best possible, due to Santolupo, was reported in [35]. The same survey article also posed the more general question of determining $\text{sat}(Q_d, Q_m)$.

Together with J. R. Johnson, in [49] we construct (Q_d, Q_m) -saturated graphs of arbitrarily low edge density, for all fixed m , thus both generalizing and strengthening the bound

of Choi and Guan. In other words, we show:

Theorem 2.1. *For fixed m ,*

$$\lim_{d \rightarrow \infty} \frac{\text{sat}(Q_d, Q_m)}{e(Q_d)} = 0.$$

Slightly more precisely, we show that $\text{sat}(Q_d, Q_m) \leq \frac{c_1}{d^{c_2}} e(Q_d)$, where c_1 and c_2 are constants depending on m . In the case $m = 2$, $c_2 = \frac{6}{7}$; it is higher for larger values of m .

We also in [49] prove a stronger bound for the semi-saturation version of the problem.

Theorem 2.2. *For all d, m , $s\text{-sat}(Q_d, Q_m) < (m^2 + \frac{m}{2})2^d$.*

We then adapt this proof in the $m = 2$ case to remove all copies of Q_2 and thus prove a bound on $\text{sat}(Q_d, Q_2)$ much stronger than that given by Theorem 2.1.

Theorem 2.3. *For all d , $\text{sat}(Q_d, Q_2) < 10 \cdot 2^d$.*

After our manuscript [49] was placed on the arxiv, Morrison, Noel and Scott [76] used similar techniques to our proof of Theorem 2.3 to prove bounded average degree bounds for (Q_d, Q_m) -saturated graphs for arbitrary m .

We briefly mention here a somewhat related saturation problem on the hypercube, although we do not work with it in this chapter. Here, Q_d is considered as $\mathcal{P}(X)$, the power set of an d element set, X . Let F be a fixed poset. A family $\mathcal{A} \subseteq \mathcal{P}(X)$ is said to be F -saturated if there is no subfamily of \mathcal{A} with the same poset structure as F , but adding any set to \mathcal{A} destroys this property. Both the maximum and minimum size of such \mathcal{A} have been studied—see for instance Katona and Tarján [58] for the former and Morrison, Noel and Scott [75] for the latter.

In [2], Alon, Krech and Szabò discuss an interesting hypergraph-type generalization of the Turán problem on the hypercube. We write Q_d^t for the 2^t -uniform hypergraph with vertex set $\{0, 1\}^d$ and edge set consisting of all t -dimensional subcubes of Q_d . We say that a subhypergraph H of Q_d^t is Q_m^t -free if it contains no subhypergraph isomorphic to Q_m^t . As in the usual ($t = 1$) case of this Turán problem, they ask how many (hyper-)edges H can have, in particular asking for the limit: $\lim_{n \rightarrow \infty} \max \left\{ \frac{e(H)}{\binom{n}{t} 2^{n-t}} \right\}$. This question is still open, but it is interesting to know that the corresponding saturation problem can be attacked by the same method as the proof of Theorem 2.1'.

We write $\text{sat}(Q_d^t, Q_m^t)$ for the smallest number of edges a (Q_n^t, Q_m^t) -saturated hypergraph can have. We show by the same method as the proof of Theorem 2.1 that, for fixed $t \geq 0$ and $s \geq 0$,

$$\lim_{d \rightarrow \infty} \frac{\text{sat}(Q_d^t, Q_{t+s}^t)}{\binom{d}{t} 2^{d-t}} = 0.$$

Another topic related to saturated graphs is the (H, F) -free process, a natural constrained random process for generating (H, F) -saturated graphs. We consider a sequence of subgraphs of H , written G_0, \dots, G_M , where G_0 has no edges and G_{i+1} is formed by adding to G_i an edge in $E(H) \setminus E(G_i)$ uniformly at random, subject to the constraint that we do not add edges that form copies of F . We terminate the sequence when there are no edges we may add, i.e. the graph G_M is (H, F) -saturated.

Equivalently, the sequence of graphs may be generated by first picking a uniformly random permutation of the edges of H . We again let G_0 be the empty graph on $|H|$ vertices. Having constructed G_i and considered the first t edges in the permutation, we form G_{i+1} by adding to G_i the next edge in the order that does not create a Q_2 .

Work in this direction was initiated in 1992 by Ruciński and Wormald [86], who studied the case where $H = K_n$ and $F = K_{1,3}$, the star with three leaves, investigating the structure of G_M .

A major breakthrough in this area was the 2009 paper of Bohman [11] on the case $H = K_n$ and $F = K_3$, the so-called triangle-free process. Using the differential equations method for random graph processes introduced in [86] (see for instance Wormald [92] for a survey of the subject), Bohman determined the order of M , the number of edges added by the process, with high probability. Indeed, he showed:

Theorem 2.4 (Bohman [11]). *Let G_M be the graph generated by the triangle-free process. Then with high probability,*

$$c_1(\log n)^{\frac{1}{2}} n^{\frac{3}{2}} \leq M \leq c_2(\log n)^{\frac{1}{2}} n^{\frac{3}{2}},$$

for some constants c_1 and c_2 .

This result was later refined by Pontiverios, Griffiths and Morris [83] and independently by Bohman and Keevash [13]. Both sets of authors used a substantial extension of the

differential equations method to determine M asymptotically, with high probability. Both sets of authors also used their analysis of the triangle-free process to improve the known bounds on the Ramsey number $R(3, t)$.

Here we consider instead the (Q_d, Q_2) -free process. Despite being a very natural variant of the triangle-free process, it appears not to have been studied much previously. Using the second moment method, we give the following lower bound on the number of edges in the process.

Theorem 2.5. *Let M be the number of edges in the (Q_d, Q_2) -free process. With high probability, $M > cd^{2/3}2^d$, for some constant, c .*

The remainder of the chapter is laid out as follows. In Section 2.2, we introduce an object we use in all of our upper bounds, the Hamming code.

In Section 2.3, we give the proof of our zero density bound on $\text{sat}(Q_d, Q_m)$, Theorem 2.1, which we consider one of the main results of this chapter. In Section 2.4, we use the same density incrementation approach to prove a generalization of this to the hypergraph setting.

In Section 2.5, we prove our bounded average degree bounds, for $s\text{-sat}(Q_d, Q_m)$ and for $\text{sat}(Q_d, Q_2)$. We consider this latter bound to be our other main result in this chapter.

It is easy to see that both Theorem 2.2 and 2.3 are best possible up to a constant factor, as all (Q_d, Q_m) -semi-saturated graphs have minimum degree $m - 1$. In Section 2.6 we improve this trivial lower bound, by showing that $s\text{-sat}(Q_d, Q_2) \geq \frac{m+1}{2} 2^d$.

We deal in Section 2.7 with the (Q_d, Q_2) -free process, proving Theorem 2.5, our lower bound on the number of edges of the (Q_d, Q_2) -free process. We use a heuristic from [11] to suggest the true order of M . However, we also note interesting differences between the (Q_d, Q_d) -free process and the triangle-free process that makes proving this heuristic using Bohman's methods problematic.

Lastly, in Section 2.8, we discuss further related questions related to saturation.

2.2 Preliminaries

An object we shall use in several of our constructions is the *Hamming code*, introduced in 1950 by Hamming [42]. The properties of Hamming codes that we require are listed below, but see van Lint [90] for more background. For our purposes, a Hamming code, C , can be thought of as a subset of $V(Q_d)$, where $d = 2^r - 1$ for some r , with the following properties:

1. C is a linear subspace of \mathbb{F}_2^d . More precisely, C is the kernel of an r by d matrix H over the field \mathbb{F}_2 , called a parity check matrix. The columns of H are precisely the non-zero vectors in \mathbb{F}_2^r .
2. $|C| = \frac{2^d}{d+1}$.
3. C has minimum distance 3. In other words, $\min\{d(x, y) : x, y \in C\} = 3$.
4. C is a dominating set for Q_d . In other words, every vertex of Q_d is either in C or adjacent to a vertex in C .

Property 1 is usually taken as the definition of a Hamming code; the other properties are simple consequences of it.

A subset C with these properties exists only if $d = 2^r - 1$. When it exists, it is the largest set with Property 3, and the smallest with Property 4. For other values of d , we use what we call an *approximate Hamming code*. This is a set $C \subset V(Q_d)$ satisfying:

1. C is a linear subspace of \mathbb{F}_2^d . More precisely, C is the kernel of an $r = \lceil \log(d+1) \rceil$ by d matrix H over the field \mathbb{F}_2 . H has as columns any d distinct non-zero binary vectors of length r .
2. $|C| = \frac{2^d}{2^{\lceil \log_2(d+1) \rceil}}$.
3. C has minimum distance 3. In other words, $\min\{d(x, y) : x, y \in C\} = 3$.

2.3 Zero Density Bound on $\text{sat}(Q_d, Q_m)$

In this section, we shall prove a quantitative version of Theorem 2.1, of which Theorem 2.1 is an immediate consequence.

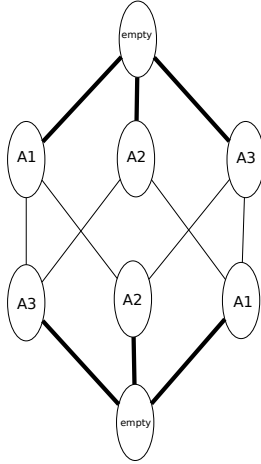


Figure 2.1: The ‘almost’ saturated graph, B

Theorem 2.1’. *For all $m \geq 1$, there exist constants, $c_m > 0$ and $a_m > 0$, such that $\text{sat}(Q_d, Q_m) \leq \frac{c_m}{n^{a_m}} e(Q_d)$. More precisely, $a_1 = 1$, and $a_m = \frac{1}{7 \cdot 3^{m-2}}$, for all $m > 1$.*

Before discussing the proof of Theorem 2.1’, we sketch a proof of the $(\frac{1}{4} + o(1))$ density bound of Choi and Guan, as this contains the main ideas of the proof of Theorem 2.1’. This proof is significantly different from Choi and Guan’s, which may be considered more direct. However, our approach, which uses $\frac{1}{3} + o(1)$ density saturated graphs to build $\frac{1}{4} + o(1)$ density saturated graphs, gives rise naturally to an iterative approach for proving Theorem 2.1’.

We assume that there exist three (Q_d, Q_2) -saturated graphs, A_1, A_2 and A_3 of $\frac{1}{3} + o(1)$ density, such that every edge of Q_d lies in one of them. Such A_i are relatively easy to construct—we will require a generalization of them in our proof of Theorem 2.1’. We now use them to produce a $\frac{1}{4} + o(1)$ density (Q_{d+3}, Q_2) -saturated graph B' .

We first construct an ‘almost’ (Q_{d+3}, Q_2) -saturated graph B . We consider Q_{d+3} as $Q_d \square Q_3$. We leave two principal Q_d ’s corresponding to antipodal vertices of Q_3 empty. Around each of these empty Q_d , we arrange copies of A_1, A_2, A_3 , as in Figure 2.1. We also add all external edges with one endpoint in either of the two empty principal Q_d ’s (as indicated by the bold edges in the figure).

The graph constructed has the property that for any edge of an empty Q_d , e , a corre-

sponding edge, e' , is present in one of the A_i . So adding e to B forms a Q_2 comprising e , e' and the two external edges that connect corresponding endpoints of e and e' . Since the A_i are themselves Q_2 -saturated graphs, adding any internal edge forms a copy of Q_2 .

It is easy to see that B is still Q_2 -free, and a simple calculation shows that B has edge density $\frac{1}{4} + o(1)$. We now prove a simple lemma that allows us to extend B to a Q_2 -saturated graph without significantly increasing the number of edges.

Lemma 2.6. *Fix $m \geq 2$. Suppose that G is a Q_m -free subgraph of Q_d and $S \subseteq E(Q_d)$. Then we can form a Q_m -free graph G' by adding no more than $|S|$ edges to G with the property that adding any edge in $S \setminus E(G)$ forms a copy of Q_m .*

Proof. We order the edges in S arbitrarily. Consider these edges in this order and add them to G if and only if doing so does not form a copy of Q_m . Since only edges of S are added by the process, we are done. \square

We apply this lemma to B , with S being the set of external edges, that have not already been added, i.e. those represented by the thin edges in Figure 1. This forms a Q_2 -saturated graph, B' . Since there are $\frac{3}{d+3}e(Q_{d+3})$ external edges, the asymptotic edge density is still $\frac{1}{4}$ and thus the graph B' attains the bound of Choi and Guan.

The proof of Theorem 2.1' uses a similar method multiple times to produce (Q_d, Q_m) -saturated graphs of arbitrarily low density. In the case where $m = 2$, we assume that we have a collection of Q_2 -saturated graphs, A_1, \dots, A_k , of edge density at most ρ , such that every edge of Q_d is contained in at least one of the A_i . We will view Q_{d+k} as $Q_d \square Q_k$ and leave several principal Q_d 's empty and fill the other principal Q_d 's with copies of some A_i . We shall ensure that each empty Q_d is adjacent, for every i , to a principal Q_d filled with A_i , and add every external edge leaving these empty Q_d 's. This ensures that adding an edge within the empty Q_d forms a copy of Q_2 . The constraint we need on the empty principal Q_d is that the set of vertices that we replace with empty Q_d 's must have minimum distance 3, and so we employ a Hamming code, enabling us to produce a graph with a lower density, ρ' . Of course, to apply this method again, we need several (Q_{d+k}, Q_m) -saturated graphs of density ρ' , which between them cover the edges of Q_{d+k} . This turns out to be not much harder, using cosets of the Hamming code.

In the general m case we adapt this method. We would like to use a collection of A_i that cover all the copies of Q_{m-1} in Q_d . Such a collection seems hard to construct, but a modification of the above argument shows that it suffices to cover almost all copies of Q_{m-1} . The other modification is that instead of using empty principal Q_d , we fill them with low density Q_{m-1} -saturated graphs, which we may assume exist by induction on m .

Proof of Theorem 2.1'. We proceed by an induction on m , wherein the inductive step uses the ‘density increment’ argument sketched above.

Base case: $m = 1$. This is trivial—the subgraph of Q_d with no edges is Q_1 -saturated.

Inductive step: take $m > 1$ and assume the Theorem holds for $m - 1$.

Claim 2.7. *Suppose we have a collection, A_1, \dots, A_k , of (Q_d, Q_m) -saturated graphs, each of density at most ρ , and some d_0 such that every Q_{m-1} that lies along the first d_0 directions is within one of these A_i . Then there is a collection of $k + 1$ (Q_{d+k}, Q_m) -saturated graphs, B_0, \dots, B_k , such that every Q_{m-1} that lies along the first d_0 directions is in one of these B_i . Further, each of the B_i has density at most $(1 - \frac{1}{2k})\rho + f(d, d_0)$, where f is a function that tends to zero whenever $d, d_0 \rightarrow \infty$ in such a way that $\frac{d_0}{d} \rightarrow 1$.*

A precise upper bound on the densities of the B_i is required for the quantitative part of the theorem; this will be stated at the end of the proof of this claim.

Proof of Claim 2.7. We start by constructing a proper $(k + 1)$ -colouring, c_0 , of Q_k , with the colours $0, 1, \dots, k$. Fix C_0 , an approximate Hamming code in Q_k . We set $c_0(x) = 0$ for all $x \in C_0$ and for all $j \in \{1, \dots, k\}$ and all $x \in C_0$, we set $c_0(x + e_j) = j$. Note that when $k + 1$ is not a power of 2 (i.e. when we do not have a genuine Hamming code), this colouring is not fully defined, since C is not dominating. For now we assign arbitrary colours other than 0 to these vertices, but we will later decide on these colours.

We write $Q_{d+k} = Q_d \square Q_k$. We induce from c_0 a colouring on the set of principal Q_d ’s in the natural way. We start forming the graph B_0 by placing a copy of A_j in each principal Q_d coloured j , for each $j \neq 0$. Also, we add to the graph B_0 every external edge with one endpoint in a principal Q_d coloured 0.

By our induction hypothesis, there exists a Q_{m-1} -saturated subgraph, G , of Q_d with no more than $\frac{c_{m-1}}{d^{a_{m-1}}} e(Q_d)$ edges. We place a graph isomorphic to G in each Q_d that is coloured

0 (we will choose which isomorphism later).

Notice that so far, B_0 is Q_m -free. Indeed, suppose that B_0 does contain a Q_m . This Q_m cannot lie entirely within a single principal Q_d , by our assumption that the A_i are saturated. As we have only added external edges that leave Q_d coloured 0, the Q_m may contain an edge between two principal Q_d 's only if one of them is coloured 0. Since the approximate Hamming code has minimum distance 3, the Q_m must contain edges in exactly two principal Q_d 's, one of which is coloured 0. But such Q_d are Q_{m-1} -saturated and thus contain no Q_{m-1} , yielding a contradiction.

So far, B_0 is not quite Q_m -saturated—for instance adding an external edge may not create a copy of Q_m . However, we use Lemma 2.6 to remedy this. We add at most $\frac{k}{d+k}e(Q_{d+k})$ edges to B_0 and we now only need to consider adding internal edges.

Adding an edge within a Q_d coloured $j \neq 0$ forms a Q_m , as each A_j is Q_m -saturated. Adding an edge within a principal Q_d coloured 0 will form a Q_{m-1} within that Q_d . If that Q_{m-1} only uses edges in the first d_0 directions, it lies within one of the A_j by the hypothesis of Claim 2.7. Since every principal Q_d coloured zero is adjacent to a principal Q_d of every non-zero colour, a Q_m will be formed. Therefore, we only need to worry about adding edges to G if the Q_{m-1} formed does not lie exclusively along the first d_0 directions—we call such edges *bad edges*. We will now show that we may assume there are not very many bad edges.

Apply a random automorphism of Q_d to G , our low density Q_{m-1} -saturated graph. We call the graph formed $G' \subseteq Q_d$, which is to be placed within a principal Q_d coloured 0. Let e be a fixed edge of this principal Q_d .

$$\begin{aligned} \mathbb{P}(e \text{ is a bad edge}) &\leq 1 - \frac{d_0}{d} \cdot \frac{d_0 - 1}{d - 1} \cdot \dots \cdot \frac{d_0 - m + 2}{d - m + 2} \\ &\leq 1 - \frac{(d_0 - m)^{m-1}}{d^{m-1}} \\ &= \frac{d^{m-1} - (d_0 - m)^{m-1}}{d^{m-1}}. \end{aligned}$$

This tells us that the expected number of bad edges, in each principal Q_d coloured 0, is no more than $\left(\frac{d^{m-1} - (d_0 - m)^{m-1}}{d^{m-1}}\right)e(Q_d)$. We now choose the automorphism of G that we left unspecified earlier; we can do this such that we get no more bad edges than the expected number. We use Lemma 2.6, with S being the set of bad edges, to form a graph that we

also call B_0 that is Q_m -saturated.

We now construct the other graphs B_i to cover the required Q_{m-1} 's. To construct B_i , we repeat the same method used for constructing B_0 , except we use $C_i := \{c + e_i : c \in C_0\}$ instead of C_0 . Note that we can make the arbitrary choices of colours to ensure each principal Q_d is filled with each of the graphs A_1, \dots, A_k , in one of the B_i .

It is easy to see that the B_i satisfy the necessary Q_{m-1} condition. Indeed any $Q_m \subseteq Q_{d+k}$ along the first d_0 directions must lie within a principal Q_d . When considered as a subgraph of this Q_d , it must lie in a copy of one of the A_i —say A_j . This principal Q_d is filled with A_j in one of the B_i , so we are done.

It remains only to bound the number of edges in each saturated subgraph, B_i . Let $e_A = \max\{e(A_i)\}$, $e_B = \max\{e(B_i)\}$, $\rho_A = \frac{e(A)}{d^{2^{d-1}}}$ and $\rho_B = \frac{e(B)}{(d+k)^{2^{d+k-1}}}$. In the calculation that follows, we write $a = a_{m-1}$ and $c = c_{m-1}$ for brevity.

Recall that edges were added to each B_j in 4 ways: from copies of A_i , from adding external edges, from the Q_{m-1} -saturated graphs and from adding bad edges.

$$\begin{aligned} e_B \leq & 2^k \left(1 - \frac{1}{2^{\lceil \log(k+1) \rceil}} \right) e_A + \frac{k}{d+k} e(Q_{d+k}) \\ & + \frac{2^k}{2^{\lceil \log(k+1) \rceil}} e(Q_d) \left(c_{m-1} d^{-a} + \frac{d^{m-1} - (d_0 - m)^{m-1}}{d^{m-1}} \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \rho_B \leq & \left(1 - \frac{1}{2^{\lceil \log(k+1) \rceil}} \right) \rho_A + \frac{k}{d+k} \\ & + \frac{1}{2^{\lceil \log(k+1) \rceil}} \left(c_{m-1} d^{-a} + \frac{d^{m-1} - (d_0 - m)^{m-1}}{d^{m-1}} \right) \\ \leq & \left(1 - \frac{1}{2k} \right) \rho(A) + \frac{k}{d} + \frac{1}{k} \left(c_{m-1} d^{-a} + \frac{d^{m-1} - (d_0 - m)^{m-1}}{d^{m-1}} \right). \end{aligned}$$

Clearly if d_0 is large enough, and $d = (1 + o(1))d_0$, the last two terms can be arbitrarily small, thus concluding the proof of the claim. □

We first find a collection of subgraphs, A_1, \dots, A_{m+1} , of Q_{d_0} that satisfy the hypothesis of Claim 2.7, with $\rho = 1$. To do this, let A_i initially consist of all edges whose lowest weight endpoint has weight in $\{i, \dots, i + m - 2\} \pmod{m+1}$, and then extend greedily until A_i

is Q_m saturated. Each A_i contains every Q_{m-1} whose lowest weight vertex has weight $i \bmod m + 1$, so every Q_{m-1} is contained in one of these A_i . Trivially, we may bound the density of these A_i above by 1, and it is easy to see this is best possible up to a constant.

We now apply Claim 2.7 repeatedly, t times. We write k_i and d_i for the value of k and d after the i^{th} iterate. Clearly, $k_0 = m + 1, k_{i+1} = k_i + 1, d_{i+1} = d_i + k_i$ and $d_t = d_0 + \sum_{i=m}^{m+t-1} i = d_0 + O(t^2)$.

After t steps, we end with saturated graphs of density, ρ :

$$\begin{aligned} \rho &\leq \prod_{i=0}^{t-1} \left(1 - \frac{1}{2k_i}\right) + \sum_{i=0}^{t-1} \left(\frac{k_i}{d_i} + \frac{c_{m-1}}{k_i} \cdot d_i^{-a} + \frac{d_i^{m-1} - (d_0 - m)^{m-1}}{k_i d_i^{m-1}}\right) \\ &\leq c \prod_{i=1}^{m+t} \left(1 - \frac{1}{2i}\right) + \frac{t(m+t)}{d_0} + \frac{tc_{m-1}}{m} \cdot d_0^{-a} + \frac{t}{m} \frac{d_t^{m-1} - (d_0 - m)^{m-1}}{d_0^{m-1}} \\ &= c' \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^{m+t} \frac{1}{i}\right) + O(t^2 d_0^{-1}) + O(t d_0^{-a}) + O\left(\frac{t^3}{d_0}\right) \\ &= O(t^{-\frac{1}{2}}) + O(t d_0^{-a}) + O(t^3 d_0^{-1}). \end{aligned}$$

Here, c and c' are constants dependent on m . If $m = 2$, we take $t = n_0^{2/7}$. Otherwise we take $t = d_0^{2a/3}$, yielding the required bound.

□

Note that the better bound for $\text{sat}(Q_d, Q_2)$ in the next section can be fed into the induction in the theorem to produce the slightly better bound of $a_m = \frac{1}{7 \cdot 3^{m-3}}$.

2.4 Saturated Hypergraphs

In this section we prove our generalization of Theorem 2.1 to the hypergraph setting. Note that we do not prove quantitative upper bounds on $\text{sat}(Q_d^t, Q_{t+s}^t)$, that is, we do not state or prove a version of Theorem 2.1', although our method allows us to do this.

Theorem 2.8. *For all fixed $t \geq 0$ and $s \geq 0$,*

$$\lim_{d \rightarrow \infty} \frac{\text{sat}(Q_d^t, Q_{t+s}^t)}{\binom{d}{t} 2^{d-t}} = 0.$$

The proof is similar to the proof of Theorem 2.1. We use induction on s and the induction step uses an density incrementation argument that is almost very similar to that used for Claim 2.7.

Proof. Base case: $s = 0$. This is trivial, since the empty subhypergraph of Q_d^t is (Q_d^t, Q_{t+s}^t) -saturated.

Induction step: We take $s \geq 1$ and assume the theorem holds for $s - 1$.

Claim 2.9. *Given a collection A_1, \dots, A_k of (Q_d^t, Q_{t+s}^t) -saturated hypergraphs, each of density at most ρ , and some d_0 such that every Q_{t+s-1}^t that lies along the first d_0 directions is contained in one of these A_i . Then there is a collection of $k + 1$ (Q_{d+k}^t, Q_{t+s}^t) -saturated graphs, B_0, \dots, B_k , such that every Q_{t+s-1}^t that lies along the first d_0 directions is in one of these B_i . Further, each of the B_i has density at most $(1 - \frac{1}{2k})\rho + f(d, d_0)$, where f is a function that tends to zero whenever $d, d_0 \rightarrow \infty$ in such a way that $\frac{d_0}{d} \rightarrow 1$.*

Proof of Claim 2.9. We view Q_{d+k} as $Q_d \square Q_k$ and we use the same approach as in the proof of Claim 2.7 to colour the principal Q_d 's with the colours $0, 1, \dots, k$. We start constructing the hypergraph G_0 by placing A_j in each Q_d coloured j , for all $j \neq 0$. We also add every t -dimensional face with exactly 2^{t-1} vertices in one of the principal Q_d 's coloured 0.

By our induction hypothesis, there exists an arbitrarily low-density Q_{t+s-1}^t -saturated subhypergraph, G , of Q_d^t . We place a hypergraph isomorphic to G in each Q_d that is coloured 0 (we will choose which isomorphism later). For d large enough, the fraction of all t -dimensional facets added in this step is arbitrarily small.

It is easy to see that B_0 is Q_{t+s}^t -free. So far, however, B_0 is not quite Q_{t+s}^t -saturated—for instance adding an external t -dimensional face that does not create a copy of Q_{t+s}^t . We greedily add every external t -dimensional face that does not create a copy of Q_{t+s}^t . If $\frac{d_0}{d}$ is sufficiently close to 1, the proportion of t -dimensional faces we add is an arbitrarily small proportion of the number of t -dimensional faces in Q_d^t .

Adding an edge within a Q_d coloured $j \neq 0$ forms a Q_{t+s} , as each A_j is Q_{t+s} -saturated. Adding an edge within a principal Q_d coloured 0 will form a Q_{t+s-1} within that Q_d . If that Q_{t+s-1} only uses edges in the first d_0 directions, it lies within one of the A_j by the hypothesis of Claim 2.7. Since every principal Q_d coloured zero is adjacent to a principal Q_d of every

non-zero colour, a Q_{t+s} will be formed. Therefore, we only need to worry about adding faces to G if the Q_{t+s-1} formed does not lie in the first d_0 directions. We call such faces *bad faces*. As in the proof of Claim 2.7, we can choose the isomorphism of G such that the proportion of t -faces that are bad is arbitrarily small, and we add such faces to B_0 if adding them does not form a copy of Q_{t+s}^t . This results in B_0 being a (Q_{d+k}^t, Q_{t+s}^t) -saturated hypergraph, of density arbitrarily close to $(1 - 2^{\lceil \log(k+1) \rceil})$.

We now construct the other B_i by repeating the same method used for constructing B_0 , using $C_i := \{c + e_i : c \in C_0\}$ instead of C_0 . Note that we may once again ensure that each principal Q_d is filled with each of the graphs A_1, \dots, A_k , in one of the B_i .

It is easy to see that these B_i satisfy the requirements of the theorem. \square

All that remains is to find a collection of subhypergraphs A_1, \dots, A_{t+s+1} of Q_{d_0} that satisfy the conditions of Claim 2.9.

We initially let A_i consist of all t -faces whose lowest weight vertex has weight in $\{i, \dots, i + t + s - 2\} \bmod m + 1$ and extend each A_i greedily until it is saturated.

Having done this, for any $\epsilon > 0$, we may choose a large enough d_0 such that applying the claim enough times yields a saturated hypergraph of density at most ϵ , concluding the proof. \square

2.5 Bounded Average Degree Constructions

2.5.1 Semi-saturation

In this section we will prove Theorem 2.2, by constructing for each m a family of Q_m -semi-saturated graphs with bounded average degree. Although it seems difficult in general to make these graphs Q_m -free, in the $m = 2$ case we will use similar ideas to prove Theorem 2.3.

In what follows it will be useful to write $d = m(2^t - 1) + r$, where $0 \leq r < m2^t$, and to let $d_0 = 2^t - 1$. We write a vertex of Q_d as $(v_1|v_2|\dots|v_m|v_{m+1})$, where $v_i \in \{0, 1\}^{d_0}$ for $i \leq m$ and $v_{m+1} \in \{0, 1\}^r$. The final section of the vector is only included to make the number of coordinates exactly d but otherwise has no importance in the construction.

Proof of Theorem 2.2. Let $C \subseteq \{0, 1\}^{d_0}$ be a Hamming Code. We define:

$$A = \{(v_1 | \dots | v_m | v_{m+1}) \in V(Q_d) : \exists i \in \{1, m\} \text{ such that } v_i \in C\}.$$

We form $E(G)$ by picking all edges with at least one endpoint in A . Note that vertices in A have degree d in G ; all other vertices have degree m . Therefore $e(G) = \frac{1}{2} ((d - m)|A| + m2^d) \leq \frac{m}{2} \left(d \frac{2^d}{(d_0+1)} + 2^d \right)$. As $\frac{d}{d_0} < 2m$, the number of edges of G satisfies the bounds of the theorem.

We now show that G is Q_m -semi-saturated. Assume $e \in E(Q_d) \setminus E(G)$ is along a direction, i , in $\{1, d_0\}$ (all other cases can be dealt with similarly). We write the endpoints of the edges as $(v_1 | v_2 | \dots | v_m | v_{m+1})$ and $(v'_1 | v_2 | \dots | v_m | v_{m+1})$, where v'_1 and all of the v_i do not lie in C . Thus for $i = 2, 3, \dots, m$ there exists $c_i \in C$ adjacent to v_i . Consider the 2^m points of the form $(x_1 | \dots | x_m | v_{m+1})$, where $x_1 \in \{v_1, v'_1\}$ and for $i = 2, 3, \dots, m$, $x_i \in \{v_i, c_i\}$. These vertices form a subcube of Q_d and all but the endpoints of e are in A . Thus when the edge e is added, a copy of Q_m is formed, concluding our proof.

□

Remark 2.10. Clearly, when $d = m(2^t - 1)$ for some t , we get the slightly stronger bound $s\text{-sat}(Q_d, Q_m) \leq \left(\frac{m^2}{2} + \frac{m}{2} \right) 2^d$.

2.5.2 Improved bound for $\text{sat}(Q_d, Q_2)$

In the $m = 2$ case, the Q_2 -semi-saturated graph constructed above consists of all edges incident with vertices in $A = \{(v_1 | v_2 | v_3) \in V(Q_d) : v_1 \in C \text{ or } v_2 \in C\}$. It is easy to see this contains large subcubes, of the form $(c | *, \dots, *, *, \dots, *)$ or $(*, \dots, * | c | *, \dots, *)$, for $c \in C$. There are other Q_2 's in this graph, but those within these large subcubes are hardest to deal with. We prevent subcubes of the first type by only adding edges of the form $\{(c | v), (c | v')\}$, where $c \in \{0, 1\}^{d_0}$ and $v \in \{0, 1\}^{d-d_0}$ and v has lower weight than v' , if v_1 has even weight. A simple parity argument shows that such edges cannot form a square. Of course doing just this alteration means the graph is no longer semi-saturated; we get around this by picking a subset D of $V(Q_{d_0})$ with similar properties to C , and adding edges starting at $(c | v_2 | v_3)$ if $(v_2 | v_3)$ contains an odd number of 1's and if $c \in D$.

Proof of Theorem 2.3. As before, we write $d = 2(2^t - 1) + r$, where $0 \leq r < 2^{t+1}$, and let $d_0 = 2^t - 1$. We write an element, x , of $\{0, 1, *\}^d$ as $(x_1|x_2|x_3)$, where $x_1, x_2 \in \{0, 1, *\}^{d_0}$ and $x_3 \in \{0, 1, *\}^r$. We refer to x_1 as the first part of x , x_2 as the second part and so on. We will use this notation particularly when x represents a vertex or an edge of Q_d (it contains no stars or one star).

Claim 2.11. *There exists a Q_2 -free spanning subgraph, H , of Q_{d_0} , that has two independent dominating sets, $C, D \subset V(H) = \{0, 1\}^{d_0}$, with C disjoint from D , where $|C| = 2^{d_0}/(d_0 + 1)$ and $|D| = 3 \cdot 2^{d_0}/(d_0 + 1)$. Further, H only contains edges incident with $C \cup D$ and $e(H) \leq 2^{d_0+1}$.*

We shall prove this claim later, first we show why it implies the theorem. We start by constructing a graph G that is Q_2 -free and will then use Lemma 2.6 to add a ‘few’ edges ($o(2^d)$ edges) to form G' , a Q_2 -saturated graph. As in the proof of Theorem 2.2, we will define a subset, A of the vertices, which will be dominating in G :

$$A = \{(v_1|v_2|v_3) \in \{0, 1\}^d : v_1 \in C \cup D \text{ or } v_2 \in C \cup D\}.$$

The definition of G is slightly more complicated. We add edges to $E(G)$ in three stages, and then delete some of these edges to ensure G is Q_2 -free.

Firstly, we add all edges, e , where $e_1 \in C$, and the remainder, $(e_2|e_3)$, contains an even number of 1’s and a single star, as well as edges where $e_2 \in C$ and the remainder, $(e_1|e_3)$ contains an even number of 1’s and a single star. We call these Type 1 edges. There are $2|C|(d - d_0)2^{d-d_0-2} \leq \frac{(d-d_0)}{2(d_0+1)}2^d$ Type 1 edges.

Similarly, we add those edges, e , where $e_1 \in D$ and the remainder, $(e_2|e_3)$ contains an odd number of 1’s and a single star, as well as edges where $e_2 \in D$ and the remainder contains an odd number of 1’s and a single star. We call these Type 2 edges. There are $2(d - d_0)|D|2^{d-d_0-2} \leq \frac{3(d-d_0)}{2(d_0+1)}2^d$ Type 2 edges.

Lastly, we add all edges, e where e_1 or e_2 is an edge of H . There are $2 \cdot 2^{d-d_0}e(H) \leq 4 \cdot 2^d$ Type 3 edges.

We now delete all edges, e , which have an endpoint, $(v_1|v_2|v_3)$ such that both v_1 and v_2 lie in $C \cup D$. Thus $e(G) \leq \left(\frac{2(d-d_0)}{d_0+1} + 4\right)2^d - \frac{d2^d}{(d_0+1)^2}$.

Suppose, for contradiction, that G contains a Q_2 . Note that as all edges of G are incident with a vertex of A , this Q_2 must contain a vertex $(v_1|v_2|v_3) \in A$, where, without loss of generality, $v_1 \in C \cup D$. Note that none of the vertices can have their second part in $C \cup D$, or there is a vertex of the Q_2 with both first and second part in $C \cup D$, impossible by our deletion step.

Let s be the number of stars of the Q_2 that are in the first part of its vector representation. If $s = 2$, all four edges are Type 3 edges, impossible as H is Q_2 -free.

If instead $s = 1$, suppose the other star is in the second part (the other case is identical). Then we may write the vertices of the Q_2 as $(v_1|v_2|v_3)$, $(v'_1|v_2|v_3)$, $(v'_1|v'_2|v_3)$ and $(v_1|v'_2|v_3)$, where $v_1 \in C \cup D$ and $v_2, v'_2 \notin C \cup D$. It is easy to see that $v'_1 \in C \cup D$. By a parity argument, v_1 and v'_1 are both in C or both in D . But this is impossible as C and D are each H -independent sets.

Finally, if $s = 0$, then we can have only Type 1 edges or only Type 2 edges (depending on whether $v_1 \in C$ or $v_1 \in D$). But this is impossible by a simple parity argument.

We now show that while G is not quite saturated, it is ‘almost’ saturated. Suppose e is a Q_d -edge not incident with A . Without loss of generality, the endpoints are $(v_1|v_2|v_3)$ and $(v'_1|v_2|v_3)$, where $v_1, v'_1, v_2, v_3 \notin C \cup D$. This is an element of $E(Q_d) \setminus E(G)$. Assume that $(v_1|v_3)$ is even, (the other case is very similar) and that v'_1 has higher weight than v_1 . Then pick $c \in C$ adjacent to v_2 . $\{(v'_1|v_2|v_3), (v'_1|c|v_3)\}$ and $\{(v_1|v_2|v_3), (v_1|c|v_3)\}$ are Type 3 edges. Also, $\{(v_1|c|v_3), (v'_1|c|v_3)\}$ is a Type 1 edge as $(x|y)$ is even. Thus a Q_2 would be formed by adding the edge.

All Q_d -edges with exactly one endpoint in A are edges of G , so we only need to consider edges where one endpoint, $(v_1|v_2|v_3)$, has v_1 and $v_2 \in C \cup D$. There are $\frac{2^d}{d}$ edges of this type, and so we may use Lemma 2.6 add them greedily to G to form a Q_2 -saturated graph, G' , which has no more edges than the bound in the theorem.

Proof of Claim. Let C be a Hamming code in Q_{d_0} . For $i = 1, \dots, d_0$, let v_i be the image of the basis vector e_i under the parity check matrix, M , of the Hamming code. We may assume that $v_1 = (1, 0, \dots, 0)$, $v_2 = (0, 1, 0, \dots, 0)$ and $v_3 = (1, 1, 0, \dots, 0)$, as every vector in \mathbb{F}_2^t occurs as a column of M . We shall construct H in four stages, and then prove that it has the required properties.

1. Add to $E(H)$ every Q_{d_0} -edge adjacent to an element of C .
2. Add to $E(H)$ every Q_{d_0} -edge of the form $\{c + e_1 + e_k, c + e_1\}$, where $c \in C$, and where $k \in [4, d_0]$ is such that v_k has a 0 in the first coordinate.
3. Add to $E(H)$ every Q_{d_0} -edge of the form $\{c + e_1 + e_k, c' + e_2\}$, where $c, c' \in C$, and where $k \in [4, d_0]$ is such that v_k has a 1 in the first coordinate and a 0 in the second coordinate.
4. Add to $E(H)$ every Q_{d_0} -edge of the form $\{c + e_1 + e_k, c' + e_3\}$, where $c, c' \in C$, and where $k \in [4, d_0]$ is such that v_k has a 1 in the first coordinate and a 1 in the second coordinate.

Since C is a Hamming code, it is an independent, dominating set and $|C| = 2^{d_0}/(d_0 + 1)$. We write $C_i = \{c + e_i : c \in C\}$; in other words, $C_i = M^{-1}(v_i)$. Let $D = C_1 \cup C_2 \cup C_3$. It is easy to see every edge of H is incident with $C \cup D$. Since the C_i are disjoint translates of C , a Hamming code, $|D| = 3 \cdot 2^{d_0}/(d_0 + 1)$.

Again using that C_1 is a translate of a Hamming code, every $x \in V(Q_{d_0}) \setminus C_1$ can be written uniquely in the form $c + e_1 + e_k$ for $c \in C$ and $k \in [1, d_0]$. The restriction $k \neq 1$ is equivalent to $x \notin C$. The restriction $k \neq 2$ is equivalent to $x \notin C_3$. This is as $M(c + e_1 + e_2) = M(c) + M(e_1) + M(e_2) = v_1 + v_2 = v_3$. Similarly, $k = 3$ if and only if $x \in C_2$. Thus steps 2, 3 and 4 ensure D is independent and dominating in H .

Notice also that each $x \notin C \cup D$ is H -adjacent to exactly 1 element in D . Hence $e(H) \leq 2|Q_{d_0}|$, as required. It remains only to show that H is Q_2 -free. Suppose not. Since we have only added edges with at least one endpoint in $C \cup D$, the Q_2 must contain two opposite vertices in $C \cup D$. Since C has minimum distance 3, and since every $x \notin C \cup D$ is adjacent to only 1 element in D , one of these vertices is in D , and one is in C . Thus the vertices of the Q_2 may be written in the form $c \in C, c + e_i, c + e_j$ and $c + e_j + e_i \in C_k$, where $i, j \in [4, d_0]$ are such that $v_i + v_j = v_k$, and $k \in \{1, 2, 3\}$. But it is impossible for all the edges of this Q_2 to lie in $e(H)$. Indeed, suppose for example that $k = 3$. Then v_i and v_j must both have 1 in the first coordinate and 1 in the second coordinate, impossible if they sum to v_k . This concludes the proof of the claim.

□

□

Remark 2.12. *Again, we get a stronger bound for some values of d ; when $d = 2(2^t - 1)$ for some t , it is easy to see that $\text{sat}(Q_d, Q_2) \leq 6 \cdot 2^d$.*

2.6 Lower Bounds

All the lower bounds in this section are for $s\text{-sat}$; easily $s\text{-sat}(Q_d, Q_m) \leq \text{sat}(Q_d, Q_m)$, so the bounds are also valid for sat .

If a graph is (Q_d, Q_m) -semi-saturated, for $m \geq 2$, it must be connected. Thus it contains a spanning tree for Q_d and so $s\text{-sat}(Q_d, Q_m) \geq 2^d - 1$. This shows that Theorems 2.2 and 2.3 are best possible up to a constant factor.

Another trivial observation improves this for $m \geq 3$: if a graph is (Q_d, Q_m) -semi-saturated, it has minimum degree $m - 1$. Thus $s\text{-sat}(Q_d, Q_m) \geq \frac{m-1}{2} 2^d$.

We do better than both trivial bounds for all m

Theorem 2.13. *If $m \geq 2$, $s\text{-sat}(Q_d, Q_m) \geq \left(\frac{m+1}{2} - o(1)\right) 2^d$.*

Proof. Let G be a (Q_d, Q_2) -semi-saturated graph with minimum degree $m - 1$; note this contains all (Q_d, Q_m) -semi-saturated graphs. We call a pair (v, e) , where $v \in V(Q_d)$ and $e \in E(Q_d) \setminus E(G)$, *good* if there is a path of length 3 in G linking the endpoints of e , that passes through v , meaning v is not a start or end vertex of the path.

Note that every non-edge of G is in at least 2 good pairs, whereas each vertex v is in at most $\binom{d(v)}{2}$ good pairs.

Therefore

$$\sum_{v \in V(Q_d)} \binom{d(v)}{2} \geq 2(e(Q_d) - e(G)).$$

Subject to fixed $\sum_v d(v)$, the left hand side is maximized when the degrees are as different as possible. But no degree can be larger than d or smaller than $m - 1$. Note that $2e(G) = \sum_v d(v)$, so we have $\frac{2e(G) - 2^d}{d-1}$ vertices of degree d in this extreme case.

So certainly

$$\begin{aligned}\frac{2e(G) - (m-1)2^d}{d-1} \binom{n}{2} &\geq d2^d - 2e(G) \\ (d+2)e(G) - d(m-1)2^{d-1} &\geq d2^d \\ e(G) &\geq \left(\frac{m+1}{2} - o(1)\right) 2^d.\end{aligned}$$

□

2.7 The (Q_d, Q_2) -free process

In this section we first prove Theorem 2.5 by finding a local condition on the order of edges in a permutation that guarantees certain edges being present in the process. We then discuss the true order of the number of edges in the process.

Proof of Theorem 2.5. We use a continuous process to generate a random permutation of the edges of Q_d . Indeed, assign to each edge e , a random variable T_e , where T_e is uniformly distributed in the interval $[0, 1]$. Then if $T_e < T_f$, we say e precedes f in our order.

Let G_M be the saturated graph yielded by following the (Q_d, Q_2) -free process on this permutation.

We say that an edge is *nicely ordered* if it is not the final edge in any Q_2 in which it lies. It is easy to see that if e is nicely ordered, then e is an edge of G_M .

We shall lower bound the number of nicely ordered edges using the second moment method. Let A_e denote the indicator random variable that is 1 if e is nicely ordered and let $A = \sum_{e \in E(Q_d)} A_e$. We can see that:

$$\begin{aligned}\mathbb{P}(A_e = 1) &= \int_0^1 (1 - \tau^3)^{d-1} d\tau \\ &\geq \int_0^{d^{-\frac{1}{3}}} (1 - \tau^3)^{d-1} d\tau \\ &\geq d^{-\frac{1}{3}} \left(1 - \frac{1}{d}\right)^{d-1},\end{aligned}$$

since the integrand is a decreasing function of τ . And so

$$\mathbb{P}(A_e = 1) \geq cd^{-\frac{1}{3}},$$

for large enough d and where c is some positive constant. Thus $\mathbb{E}(A) = (1 + o(1))d^{2/3}2^d$.

It is easy to see that A_e is independent from all but $3d$ variables T_f and hence at most $9d^2$ random variables A_f . Hence,

$$\begin{aligned} \text{Var}(A) &\leq \mathbb{E}(A) + \sum_{e,f} \text{Cov}(A_e, A_f) \\ &\leq \mathbb{E}(A) + 9d^2 2^d \\ &= o(\mathbb{E}(A)^2). \end{aligned}$$

Thus by Chebychev's inequality, $A \geq cd^{2/3}2^d$ with high probability, for some c , which concludes the proof, since $M \geq A$. \square

Remark 2.14. *A slightly more careful calculation gives $\mathbb{E}(A) = \Theta(d^{2/3}2^d)$ and hence $A = \Theta(d^{2/3}2^d)$ with high probability. However, this does not give us an upper bound on M .*

Bohman [11] introduces a heuristic that assumes certain random variables related to the triangle-free process closely follow some trajectories. Using this assumption he deduces the values of those trajectories in order to bound the value of M . Using martingales, he is able to make this process rigorous. We use the same heuristic for the (Q_d, Q_2) -free process to suggest a possible order for M . However, we also point out differences between the (Q_d, Q_2) -free process and the triangle-free process that cause difficulties in making this argument rigorous.

Let G_1, \dots, G_M be the sequence of graphs generated by the (Q_d, Q_2) -free process. Let u and v be a pair of vertices that are adjacent in Q_d . We say that uv is open in G_i if there is no path of three G_i -edges that connect u to v . In other words uv is open if adding it to G_i does not form a copy of Q_2 . We write O_i for the number of open pairs in G_i . This definition of open pairs is analogous to a definition in [11].

We also define, for each Q_d -adjacent pair of vertices u and v , three other sets. Let $W_i(uv)$ denote the number of paths of length 3 from u to v with 3 open pairs in G_i , let $X_i(uv)$ be the number of paths of length 3 from u to v with 2 open pairs and one edge and let $Y_i(uv)$ count the paths of length 3 from u to v with 1 open pair and 2 edges.

For convenience, we also introduce a scaling $t = \frac{i}{d^{2/3}2^d}$. We assume there are continuous

functions q, w, x and y such that for all i and all Q_d -adjacent u and v :

$$O_i \approx q(t)d2^d, \quad W_i(uv) \approx w(t)d, \quad X_i(uv) \approx x(t)d^{2/3}, \quad Y_i(uv) \approx y(t)d^{1/3}.$$

Note that adding a single edge uv to G_i to form G_{i+1} removes $Y_i(uv)$ open edges. Thus for small ϵ , we expect

$$q(t+\epsilon)d2^d \approx O_{i+\epsilon d^{2/3}2^d} \approx O_i - \epsilon d^{2/3}2^d \cdot y(t+\epsilon)d^{1/3} \approx (q(t) - \epsilon y(t))d2^d.$$

This suggests that $\frac{dq}{dt} = -y$. Similar arguments give:

$$\frac{dx}{dt} = \frac{3w}{q} - \frac{2xy}{q}, \quad \frac{dy}{dt} = \frac{2x}{q} - \frac{y^2}{q}, \quad \frac{dw}{dt} = \frac{-3yw}{q}.$$

Solving these equations gives $q(t) = \frac{1}{2}e^{-8t^3}$, $w(t) = e^{-24t^3}$, $x(t) = 6te^{-16t^3}$ and $y(t) = 12t^2e^{-8t^3}$.

If indeed $O(i) \approx \frac{1}{2}e^{-8t^3}d2^d$, then the process ends when $t = \Theta(\log^{1/3} d)$, that is, when $i = \Theta\left((\log d)^{\frac{1}{3}}d^{\frac{2}{3}}2^d\right)$.

Due to this heuristic, we propose the following,

Question 2.1. *Let G_M be the graph generated by the (Q_d, Q_2) -free process. Is it true that with high probability,*

$$c_1(\log d)^{1/3}d^{2/3}2^d \leq M \leq c_2(\log d)^{1/3}d^{2/3}2^d,$$

for some constants c_1 and c_2 .

Associated with the (Q_d, Q_2) -free process, we have a natural sequence of graphs, $H(j)$, for $j = 0, \dots, n2^{n-1}$, where $H(j)$ is the graph formed by the first j edges in the randomly chosen permutation. The graph $H(j)$ is a natural analog of the Erdős-Renyi random graph with m edges, $G(n, m)$.

We again label the graphs of the process G_i , for $i = 0, \dots, M$, but consider i as a function of j . For Q_d -adjacent vertices u and v , note that $Y_{uv}(i(j)) = 0$ whenever u and v are isolated

in $H(j)$. Thus,

$$\begin{aligned}
\mathbb{P}(Y_{uv}(i(j)) = 0) &\geq \frac{\binom{d2^{d-1}-2d}{j}}{\binom{d2^{d-1}}{j}} \\
&= \frac{(d2^{d-1}-j) \cdots (d2^{d-1}-2d+1-j)}{(d2^{d-1}) \cdots (d2^{d-1}-2d+1)} \\
&\geq \left(1 - \frac{j}{d2^{d-1}}\right)^{2d-1} \\
&\geq \exp\left(-\frac{j}{2^{d-2}}\right).
\end{aligned}$$

There is some constant c such that while $j \leq cd2^{d-1}$, we have, in expectation, a large number of pairs uv with $Y_{uv}(i(j)) = 0$. It seems likely that i is approximately concave as a function of j , and hence it seems likely that, for some uv , the random variable $Y_{uv}(i(j))$ equals zero for a constant proportion of the process. Thus, unlike in the triangle-free process, we do not expect every variable to follow its expected trajectory closely. It is still possible that this approach can be salvaged, for instance by showing that almost every variable follows its trajectory closely, but this does not appear straightforward.

2.8 Discussion

2.8.1 (Q_d, F) -saturation

Our work raises naturally the question of the behaviour of $\text{sat}(Q_d, F)$ for general F .

For fixed F , the order of $\text{sat}(K_n, F)$ was shown to be at most linear in n by Kászonyi and Tuza in 1986 [56]. We propose the following analogue.

Conjecture 2.1. *Let F be a subgraph of Q_d , for some d , then $\text{sat}(Q_d, F) \leq c_F 2^d$, for some constant c_F depending only on F .*

As mentioned above, Morrison, Noel and Scott, in [76] show that $\text{sat}(Q_d, F) = O(2^d)$. This demonstrates that in the special case $F = Q_m$, the conjecture is true.

The special case where F is a star is trivial. Also easy is the case of a path with k edges, P_k . Indeed, if m is the largest integer such that $2^m \leq k$, the subgraph of Q_d whose edge set is the set of all edges in the first m directions is (Q_d, F) -saturated, and so $\text{sat}(Q_d, F) \leq m2^d$.

Results of Gandhi and Kim [38] show that various classes of trees, including caterpillars and generalised stars, also satisfy the conjecture.

2.8.2 Semi-saturation and weak-saturation

As mentioned in the introduction, the saturation and semi-saturation numbers may differ significantly. However, they are often equal. Indeed, $s\text{-sat}(K_n, K_t) = \text{sat}(K_n, K_t)$, for all n and t , as shown in [33].

Question 2.2. *Is $\text{sat}(Q_d, Q_2) = s\text{-sat}(Q_d, Q_2)$ for all d ? Does equality hold for all sufficiently large d ? If not, is $\liminf \frac{\text{sat}(Q_d, Q_2)}{2^d} > \limsup \frac{s\text{-sat}(Q_d, Q_2)}{2^d}$?*

Recall that all our lower bounds are for $s\text{-sat}$ —it seems hard to bound sat more strongly.

Another version of sat that has been studied in the literature (see Section 10 of [35]) (where the host graph is K_d) could be studied for this problem. We say that a graph $G \subseteq Q_d$ is (Q_d, Q_m) -*weakly-saturated* if we can add the edges in $E(Q_d) \setminus E(G)$ one at a time (in some order) such that every new edge creates at least one new copy of F . We write $w\text{-sat}(Q_d, Q_m)$ for the minimum number of edges a (Q_d, Q_m) -weakly saturated graph can have. Clearly, $w\text{-sat}(Q_d, Q_m) \leq s\text{-sat}(Q_d, Q_m) \leq \text{sat}(Q_d, Q_m)$. It is not hard to see, by induction on d , that there are many (Q_d, Q_2) -weakly-saturated trees and so $w\text{-sat}(Q_d, Q_2) = 2^d - 1$. Indeed, given any G_1, G_2 , possibly different weakly (Q_{d-1}, Q_2) -saturated trees, we place them in complementary Q_{d-1} 's, and connect any one pair of corresponding vertices. This forms a weakly (Q_d, Q_2) -saturated tree. However, $w\text{-sat}(Q_d, Q_m)$ is in general not known. Together with J. R. Johnson, we posed in [49] the following:

Question 2.3. *For $m \geq 3$, what is $w\text{-sat}(Q_d, Q_m)$?*

This was then answered by Morrison, Noel and Scott in [76], using a linear algebraic tool developed for bootstrap percolation by Balogh, Bollobas, Morris and Riordan [7].

Indeed they show:

Theorem 2.15 (Morrison, Noel and Scott [76]). *For $d \geq m \geq 1$,*

$$w\text{-sat}(Q_d, Q_m) = (m-1)2^d - \sum_{j=0}^{m-2} (m-1-j) \binom{d}{j}.$$

For fixed m , we see that this is $(m-1+o(1))2^d$.

2.8.3 Vertex saturation

A much studied variant of the Turán problem on the hypercube is the vertex Turán problem. Let F be a graph. We say a set of vertices, $S \subseteq V(Q_d)$ is F -free if the subgraph of Q_d induced by S does not contain a copy of F . We say $S \subseteq V(Q_d)$ is (Q_d, F) -saturated if it is F -free but adding any vertex to S forms a copy of F .

We let $v\text{-ex}(Q_d, F)$ denote the maximum number of vertices in a (Q_d, F) -saturated set. The function $v\text{-ex}(Q_d, F)$ has been studied by a number of authors, see for instance the work of Kostochka [63] or of Johnson and Talbot [50].

Despite this, the natural vertex saturation version of this problem seems not to have garnered much attention. We let $v\text{-sat}(Q_d, F)$ denote the minimum number of vertices in a (Q_d, F) -saturated set. We propose the following:

Question 2.4. *Suppose $m \geq 2$. What is the order of $v\text{-sat}(Q_d, Q_m)$?*

Note that this is the $t = 0$ case of hypergraph saturation, $\text{sat}(Q_d^0, Q_m^0)$. Thus Theorem 2.8 implies that $v\text{-sat}(Q_d, Q_m) = o(2^n)$, for all fixed m .

There is also a connection to notions in computer science. A t -covering code is a set $S \subseteq V(Q_d)$ for which all $v \in V(Q_d)$ are within distance t of some element of S . Suppose S is an m -covering code with minimum distance $2m - 1$. Let T denote the set of vertices within distance at most $m - 1$ from a vertex in S . It is easy to see that T is (Q_d, Q_m) -saturated, so this provides a potential method for constructing (Q_d, Q_m) -saturated sets of vertices.

Chapter 3

Directed Paths in the Cube

3.1 Introduction

In this chapter, we deal with edge-disjoint and vertex-disjoint paths in graphs, specifically the hypercube and the grid. These are much studied and fundamental structures in graphs and their presence is linked to notions of graph connectivity. The relationship between paths and connectivity was established by many results such as the Max-Flow-Min-Cut theorem and Menger's theorems, which we will make use of here. Since each of these theorems has numerous variants, we will state in Section 3.2 the versions we require, and also in Section 3.2 introduce related terminology used in the remainder of this chapter with no further introduction. A fuller introduction to the topic of flows and connectivity can be found in Chapter 3 of [16], Chapter 5 of [43] or Chapters 3.3 and 6.2 of [28].

Our work is motivated primarily by two conjectures of Bollobás and Leader on edge-disjoint and vertex-disjoint paths in the directed hypercube, in which edges are oriented towards their endpoint of highest weight. We prove strengthenings of both the conjectures and then generalize each to the grid graph. The remainder of this introduction is split into two subsections, dealing separately with edge-disjoint paths and vertex-disjoint paths.

3.1.1 Edge-Disjoint Paths

For a graph or digraph, G , we write $p_e(G, A, B)$ for the size of the largest collection of edge-disjoint paths between two disjoint subsets, A and B , of $V(G)$. We will be particularly interested in the case where G is the hypercube, Q_d , or the directed hypercube, \vec{Q}_d , and also when G is the grid or directed grid. In 1997, Bollobás and Leader [17], gave a lower bound on $p_e(Q_d, A, B)$, in terms of $|A|$ and $|B|$.

The *edge boundary* of a set S of vertices of a graph, written $\partial_e(S)$, is the set of edges with exactly one endpoint in S . There is a similar notion for directed graphs: the *directed edge boundary*, written $\vec{\partial}_e(S)$, is the set of directed edges with exactly one vertex in S , that are directed away from that vertex.

Bollobás and Leader made use of the Edge Isoperimetric Inequality which bounds the edge boundary of set families. To state the inequality, we must also define the *binary order*: for $x, y \in \mathcal{P}[d]$, we let $x <_b y$ if $\max(x \triangle y) \in y$. We write I_t for the initial segment of binary

order of size t . Thus for all k , the subcube $\mathcal{P}[k]$ is the initial segment of size 2^k , i.e. I_{2^k} . The Edge Isoperimetric Inequality, proved independently by Harper [44], Lindsey [67], Bernstein [9] and Hart [46], states that initial segments of the binary order minimize the size of the edge boundary.

Theorem 3.1 (Edge Isoperimetric Inequality). *Let $A \subseteq \mathcal{P}[d]$ with $|A| = a$. Then $|\partial_e(A)| \geq |\partial_e(I_a)|$. In particular, if $|A| = 2^k$, then its edge boundary is larger than that of a k -dimensional subcube; i.e. $|\partial_e(A)| \geq (d - k)2^k$.*

Bollobás and Leader [17] used versions of the Max-Flow-Min-Cut Theorem (see Subsection 3.2) to demonstrate a relationship between edge-disjoint paths in the hypercube and edge-boundaries of subsets, and implicitly showed a directed version of this. A *down set* is a subset, A , of Q_d such that if $x \in A$ and $y \subseteq x$, then $x \in Q_d$. An *up set* is the complement of a down set. More precisely they showed:

Lemma 3.2. *For all disjoint non-empty subsets A and B of $\mathcal{P}[d]$, we have that $p_e(Q_d, A, B) = \min\{|\partial_e(S)| : A \subseteq S \subseteq B^c\}$. If additionally A is a down set and B is an up set then $p_e(\vec{Q}_d, A, B) = \min\{\vec{\partial}_e(S) : A \subseteq S \subseteq B^c\}$.*

We give Bollobás and Leader's proof of this lemma in Section 3.1, which is based heavily on Menger's Theorem, as we will make use of it for our results. Lemma 3.2 allows us to see easily that, given the sizes of A and B , the exact lower bound for $p_e(Q_d, A, B)$ is $\min\{|\partial_e(I_t)| : |A| \leq t \leq 2^d - |B|\}$. Clearly, this lower bound is not necessarily equal to $\min\{|\partial_e(I_{|A|})|, |\partial_e(I_{|B|})|\}$, since $|\partial_e(I_t)|$ is not monotone, or even unimodal, in t . Indeed, it is easy to see that $|\partial_e(I_{2^k-1})| = (d - k)2^k + 2k - d > |\partial_e(I_{2^k})|$, when $d < 2k$.

Since $\min\{|\partial_e(I_t)| : |A| \leq t \leq 2^d - |B|\}$ is a very awkward function to work with, Bollobás and Leader provide an approximation.

Theorem 3.3 (Bollobás-Leader [17]). *Let A and B be disjoint subsets of $\mathcal{P}[d]$, each of size 2^k , for some non-negative integer k . Then there is a family of at least $(d - k)2^k$ edge-disjoint directed paths from A to B . In other words, $p_e(Q_d, A, B) \geq (d - k)2^k$.*

It is easy to see that this is best possible. Indeed, $p_e(Q_d, A, B)$ is bounded above by $|\partial_e(A)|$ and when A is a k -dimensional subcube, this is precisely $(d - k)2^k$.

Theorem 3.3 is a special case of Bollobás and Leader’s full result, which gives a lower bound for each pair of values of $|A|$ and $|B|$. The full result claimed in [17] was stated slightly incorrectly—for reasons of completeness, in Section 3.4, we give a full statement of the amended result as well as a full proof, although it is almost identical to that given by Bollobás and Leader. For now, we write $BL_e(|A|, |B|)$ for the lower bound for $p_e(Q_d, A, B)$ given by the amended result.

A *directed path* is a path in the directed hypercube. Bollobás and Leader [17] asked if the same bounds apply to edge-disjoint directed paths between A and B when A is a down set and B is an up set. More formally, they proposed the following.

Conjecture 3.1 (Bollobás-Leader [17]). *Let A , a down set, and B , an up set, be disjoint non-empty subsets of $\mathcal{P}[d]$. Then $p_e(\vec{Q}_d, A, B) \geq BL_e(|A|, |B|)$. In particular, if $|A| = |B| = 2^k$, then $p_e(\vec{Q}_d, A, B) \geq (d - k)2^k$.*

See also [22] for a brief description of their conjecture, submitted as an open problem to the 16th British Combinatorial Conference, in 2005.

In Section 3.3, we prove a strengthened version of the conjecture, that is essentially best possible:

Theorem 3.4. *Suppose A and B are disjoint subsets of Q_d , where A is a down set, and B is an up set. Then there are the same number of edge-disjoint paths from A to B as edge-disjoint directed paths, i.e. $p_e(\vec{Q}_d, A, B) = p_e(Q_d, A, B)$.*

Easily, Lemma 3.2 allows us to deduce Theorem 3.4 from the following directed version of the Edge Isoperimetric Inequality.

Theorem 3.5. *Let A be an up set and B be a disjoint down set, both non-empty subsets of Q_d . Then $\min \left\{ |\vec{\partial}_e(S)| : A \subseteq S \subseteq B^c \right\}$ is attained by a down set. Thus $\min \left\{ |\vec{\partial}_e(S)| : A \subseteq S \subseteq B^c \right\} = \min \{ |\partial_e(S)| : A \subseteq S \subseteq B^c \}$.*

We prove this Theorem in Section 3.3 using an unusual compression argument. Roughly speaking, we define two different compression operators, neither of which always reduces the size of the directed edge boundary of a set, but we show that for each set at least one of them does. The formal statement of this is the key Lemma 3.19.

In Section 3.5, we use a mild modification of our compressions to prove a version of Theorem 3.5 for the grid graph. In the context of grids, a *down set* is a subset, A , of $[m]^d$ such that if $(x_1, x_2, \dots, x_d) \in A$ and $y_i \leq x_i$ for all i , then $(y_1, y_2, \dots, y_d) \in A$. An *up set* is the complement of a down set. Further, we say $x = (x_1, x_2, \dots, x_d)$ is larger than $y = (y_1, y_2, \dots, y_d)$ if $x \neq y$ and $y_i \leq x_i$ for all i . We show that:

Theorem 3.6. *Suppose A and B are disjoint subsets of $[m]^d$, where A is a down set, and B is an up set. Then there are the same number of edge-disjoint paths from A to B as edge-disjoint directed paths, i.e. $p_e(\overrightarrow{P_m^d}, A, B) = p_e(P_m^d, A, B)$.*

3.1.2 Vertex-Disjoint Paths

For the problem of vertex-disjoint paths, in [17], Bollobás and Leader proceed in an analogous manner to the edge-disjoint case. Indeed, they use a flow theorem to establish a relationship between vertex-disjoint paths and vertex boundaries, and then apply a well-known lower bound on vertex boundaries, the Vertex Isoperimetric Inequality, to prove an analogue of Theorem 3.3.

The *vertex boundary* of a set S of vertices of a graph, G , written $\partial_v(S)$, is the set of vertices in $V(G) \setminus S$ adjacent to some vertex in S . In other words, $\partial_v(S) = \{x \in V(G) \setminus S : d(x, y) = 1, \text{ for some } y \in S\}$, where $d(x, y)$ is the usual graph distance. We note that the vertex boundary is often termed the neighbourhood of a set of vertices, and sometimes written $\Gamma(S)$. Similarly, in directed graphs, \vec{G} , the *directed vertex boundary* of $S \subseteq V(\vec{G})$, written $\vec{\partial}_v(S)$, is the set $\{v \in V(\vec{G}) \setminus S : \vec{uv} \in E(\vec{G}), \text{ for some } u \in S\}$.

The *simplicial order* on $\mathcal{P}[d]$ is defined by letting $x <_s y$ if either $|x| < |y|$ or if both $|x| = |y|$ and x precedes y in the lexicographic order, i.e. $\min(x \triangle y) \in x$. We write J_t for the initial segment of simplicial order of size t . Note that for all k , the set $[d]^{(\leq k)} := \{x \in \mathcal{P}[d] : |x| \leq k\}$ is an initial segment of simplicial order. The following theorem of Harper [45] shows that initial segments minimize vertex boundaries.

Theorem 3.7 (Vertex Isoperimetric Inequality). *Let $A \subseteq \mathcal{P}[d]$, with $|A| = a$. Then $|\partial_v(A)| \geq |\partial_v(J_a)|$. In particular, if $|A| = \sum_{i=0}^k \binom{d}{i}$, then $|\partial_v(A)| \geq \binom{d}{k+1}$.*

For a graph G , we write $p_v(G, A, B)$ for the size of the largest collection of paths with

vertex-disjoint interiors, between two disjoint subsets, A and B , of $V(G)$. Similarly, for a directed graph, \vec{G} , we write $p_v(\vec{G}, A, B)$ for the size of the largest collection of directed paths between A and B that have vertex-disjoint interiors. The link between p_v and ∂_v is a little less neat than in the edge case, as seen by the following observation from [17].

Observation 3.8. *The number of vertex-disjoint paths in Q_d between sets of vertices, A and B , is equal to $e(A, B)$ plus the smallest vertex cut separating A from B in the graph $Q_d - E(A, B)$, i.e. the graph formed by deleting all edges from A to B from the hypercube.*

Bollobás and Leader used this to give a lower bound on $p_v(Q_d, A, B)$. Their full theorem is given and discussed in Section 3.4, below is the special case in which Bollobás and Leader were most interested.

Theorem 3.9 (Bollobás-Leader [17]). *Let A and B be disjoint non-empty subsets of Q_d , with $|A| = |B| = \sum_{i=0}^k \binom{d}{k}$. Then $p_v(Q_d, A, B) \geq \binom{d}{k+1}$.*

It is easy to see that this is essentially best possible, since if A and B are non-adjacent, every path from A to B must have one vertex in $\partial_v(A)$.

We write $BL_v(|A|, |B|)$, for the lower bound given by Bollobás and Leader for $p_v(Q_d, A, B)$. As in the edge case, this lower bound is not simply the isoperimetric bound—i.e. it is not the minimum of $|\partial_v(J_{|A|})|$ and $|\partial_v(J_{|B|})|$. Indeed, they show that this minimum is not a lower bound for $p_v(Q_d, A, B)$, essentially again due to the lack of monotonicity of $|\partial_v(J_t)|$.

In [17], the proof given of Theorem 3.9 contains a small error in calculation, although the theorem is correct as stated. Essentially the same mistake is made in another theorem in [17]. We give amended proofs of both these theorems in Section 3.4.2, but we stress that the errors are minor and the majority of the proof of both statements is the same as in [17].

Bollobás and Leader also proposed a directed version of Theorem 3.9, conjecturing that their bounds hold even for directed paths between up sets and down sets:

Conjecture 3.2 (Bollobás-Leader [17]). *Let A and B be disjoint non-empty subsets of Q_d . Then $p_v(\vec{Q}_d, A, B) \geq BL_v(|A|, |B|)$. In particular, if $|A| = |B| = \sum_{i=0}^k \binom{d}{k}$, then $p_v(\vec{Q}_d, A, B) \geq \binom{d}{k+1}$.*

In Section 3.3 of this paper, we prove a strengthening of this conjecture.

Theorem 3.10. *Suppose A , a down set, and B , an up set, are disjoint non-empty subsets of Q_d . Then $p_v(\overrightarrow{Q_d}, A, B) = p_v(Q_d, A, B)$.*

As an intermediate step in the proof, we prove the following isoperimetric-type inequality, which may be of independent interest:

Theorem 3.11. *Let A be a down set and B be a up set, both non-empty subsets of Q_d . Suppose $A \subseteq S \subseteq B^c$, then there exists a down set S' satisfying $A \subseteq S' \subseteq B^c$ with $\overrightarrow{\partial}_v(S) \geq \overrightarrow{\partial}_v(S')$.*

We use a flow theorem to show a directed version of Observation 3.8, which we use to deduce Theorem 3.10 from Theorem 3.11.

It is interesting to note that although the Edge Isoperimetric Inequality and the Vertex Isoperimetric Inequality use different approaches, our two directed versions have a very similar proof, both relying on the same compressions. Again, neither of these compressions works on its own, but we show at least one of them works for each set.

Once again, we to extend this to grids, proving the natural analogue of Theorem 3.10 in Section 3.5.

Theorem 3.12. *Suppose A and B are disjoint subsets of $[m]^d$, where A is a down set, and B is an up set. Then there are the same number of vertex-disjoint paths from A to B as vertex-disjoint directed paths, i.e. $p_v(\overrightarrow{P_m^d}, A, B) = p_v(P_m^d, A, B)$.*

3.2 Flow Theorems

We state in this section the Max-Flow-Min-Cut Theorem and use it to derive the precise statements of theorems on edge-disjoint and vertex-disjoint paths that we require in the remainder of this chapter. We also prove a theorem on matchings that we require in Section 3.4.

If v is a vertex of a directed graph, we write $\Gamma^-(v)$ for its in-neighbourhood, i.e. the set of vertices u such that \overrightarrow{uv} is a directed edge and we write $\delta^-(v)$ for $|\Gamma^-(v)|$, its in-degree. We similarly define $\Gamma^+(v)$ and $\delta^+(v)$ to be the out-neighbourhood of v and out-degree of v respectively. A *flow network* is a directed graph $G = (V, \overrightarrow{E})$ with a *source* vertex s and a

sink vertex t , such that $\delta^-(s) = \delta^+(t) = 0$. A *capacity* is a function $c : \vec{E} \rightarrow \mathbb{R}_{\geq 0}$. A *cut* is a set of directed edges such that removing them from the graph removes any directed path from s to t .

A *flow* is a map $f : \vec{E} \rightarrow \mathbb{R}_{\geq 0}$ such that $f(e) \leq c(e)$, for all $e \in \vec{E}$, and such that for all $v \in V(G) \setminus \{s, t\}$,

$$\sum_{u \in \Gamma^-(v)} f(\vec{uv}) = \sum_{u \in \Gamma^+(v)} f(\vec{vu}).$$

The *value* of a flow is $\sum_{u \in \Gamma^+(s)} f(\vec{su})$. The following theorem, due to Ford and Fulkerson [37] and also stated and proved in Chapter 3 of [16], gives an effective way of calculating maximum flow values.

Theorem 3.13 (Max-Flow-Min-Cut Theorem [37]). *In a flow network with given capacity, the maximum value of a flow is equal to the minimum total capacity of the edges in a cut. Furthermore, if all of the capacities are integers, then there is a maximum value flow that is integer valued.*

We use this to prove the following variants of Menger's Theorem [72]. These theorems are naturally stated for multigraphs and directed multigraphs; that is, we allow arbitrarily many edges or directed edges between a pair of vertices.

Theorem 3.14 (Edge Menger's Theorem). *(a) Let \vec{G} be a directed multigraph, and let A and B be disjoint subsets of $V(\vec{G})$. Then $p_e(\vec{G}, A, B)$, the number of edge-disjoint directed paths between A and B , is equal to the minimum number of edges in a cut that separates A from B .*

(b) Let G be an undirected multigraph and A and B be disjoint subsets of $V(G)$. Then $p_e(G, A, B)$ is equal to the minimum number of edges in a cut that separates A from B .

Proof. (a) We form a flow network H with vertex set $V(H) = V(\vec{G}) \cup \{s, t\}$ and adding directed edges \vec{sa} and \vec{bt} of capacity $e(\vec{G}) + 1$ for all $a \in A$ and $b \in B$. For all pairs of vertices $x, y \in V(H) \setminus \{s, t\}$, we add the edge \vec{xy} with capacity equal to the number of edges starting at x and ending at y in \vec{G} .

Note that a minimum capacity cut in H corresponds to a cut separating A and B in \vec{G} ,

and hence has total capacity of the minimum cut in H equals to the minimum number of edges separating A from B in \vec{G} .

Since we have a maximum value flow that is integer-valued, it yields a union of edge-disjoint directed paths from A to B and thus Theorem 3.13 allows us to conclude.

(b) This is a simple deduction from (a). We form the directed multigraph \vec{G} by replacing each edge in G with two directed edges, in either direction. Note that the size of a minimum cut between A and B in G is the same as that in \vec{G} . Thus applying the result of (a) to \vec{G} gives the result. \square

We require also the following, again we may prove it for multigraphs, but this is unnecessary for our purposes.

Theorem 3.15 (Vertex Menger's Theorem). *Let A and B be disjoint sets of vertices in a graph G or digraph \vec{G} , such that no vertex in A is adjacent to a vertex in B .*

(a) *The number of paths in \vec{G} between A and B with vertex-disjoint interiors, $p_v(\vec{G}, A, B)$, is equal to the minimum number of vertices in a vertex cut that separates A from B .*

(b) *Similarly, $p_v(G, A, B)$ is equal to the minimum number vertices in a vertex cut that separates A from B .*

Proof. (a) We form the directed multigraph \vec{H} , with twice as many vertices as \vec{G} , labelling these as v_1 and v_2 for $v \in \vec{G}$. For each directed edge \vec{xy} of \vec{G} , we add $\vec{x_2y_1}$ to \vec{H} , with multiplicity $|\vec{G}| + 1$, and we also add for all $v \in \vec{G}$ the directed edge $\vec{v_1v_2}$. We see that a minimum edge cut in \vec{H} corresponds to a minimum vertex cut in \vec{G} , and edge-disjoint directed paths in \vec{H} correspond to vertex-disjoint paths in \vec{G} . Hence, our result follows from Theorem 3.14(a).

(b) As in the deduction of Theorem 3.14(b) from Theorem 3.14(a), we replace each edge with two directed edges and apply part (a) of this theorem to that directed graph. \square

Given a graph G , a subgraph is said to be a *matching of size k* if it has $2k$ vertices, all of degree 1. In other words a matching is collection of k independent edges of a graph.

In 1935, Hall [41] proved a powerful necessary and sufficient condition for the existence of a matching in a bipartite graph in terms of sizes of vertex boundaries. His result is often

known as Hall's Marriage Theorem. In Section 3.4, we will use the following easy corollary, which is Corollary 9 in Chapter 3 of [16], of Hall's Marriage Theorem.

Theorem 3.16 (Defect Hall's Theorem). *Let G be a bipartite graph with vertex classes U and V , such that for all $S \subseteq U$, $\partial_v(S) \geq |S| - d$. Then G contains a matching of size $|U| - d$.*

3.3 Directed Isoperimetric Inequalities and Directed Paths

3.3.1 Compression Operators

We introduce here the two different classes of compression, which we use to prove Theorems 3.5 and 3.11. Each of these compressions makes S more like a down set, in some sense that we will make concrete. For $S \subseteq \mathcal{P}[d]$, and $i \in [d]$, we say that:

$$C_i(S) = \{x \in S : x \setminus \{i\} \in S\} \text{ and } D_i(S) = S \cup \{x : x \cup \{i\} \in S\}.$$

We first state some properties of these compressions that will be used to prove both edge and vertex versions of our theorems.

Observation 3.17. *If A is a down set, B is an up set, and $A \subseteq S \subseteq B^c$, then $A \subseteq C_i(S) \subseteq S \subseteq D_i(S) \subseteq B^c$, for all i .*

We say that a set S is i -down if $x \in S \Rightarrow x \setminus i \in S$. Clearly S is a down set if and only if S is i -down for all i . It is easy to see that both $C_i(S)$ and $D_i(S)$ are i -down sets. The following lemma shows that the operators C_i and D_i preserve the j -down property.

Lemma 3.18. *Let $S \subseteq Q_d$ and $i, j \in [d]$. If S is j -down then so is $C_i(S)$ and $D_i(S)$.*

Proof. If $i = j$, this is trivial, so we assume otherwise.

Suppose $x \in C_i(S)$, then we must have $x \setminus i \in C_i(S)$, as $C_i(S)$ is i -down. Since $C_i(S) \subseteq S$, we get that x and $x \setminus i$ are in S . By our assumption that S is j -down, this implies $x \setminus j$ and $x \setminus \{i, j\} \in S$. The definition of C_i allows us to conclude that $x \setminus j \in C_i(S)$, as required.

Suppose now that $x \in D_i(S)$. Suppose first that $x \in S$ then $x \setminus j \in S$, since S is j -down. This implies that $x \setminus j \in D_i(S)$, as $S \subseteq D_i(S)$. If instead $x \notin S$, then $x \cup i \in S$, by the definition of D_i . This implies that $(x \cup i) \setminus j \in S$ and thus $x \setminus j \in D_i(S)$, again by the definition of D_i . \square

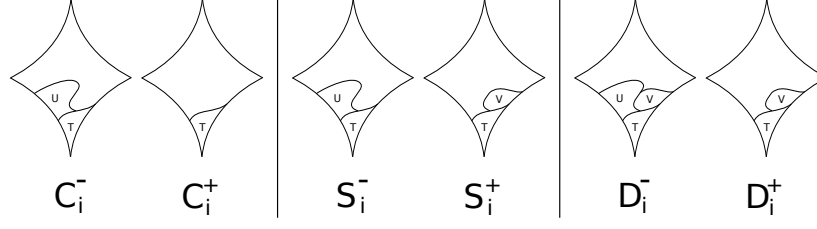


Figure 3.1: A visualization of $C = C_i(S)$ and $D = D_i(S)$

For $S \subseteq \mathcal{P}[d]$, the i -sections of S are the sets $S_i^+ := \{x \in \mathcal{P}([d] \setminus \{i\}) : x \cup \{i\} \in S\}$ and $S_i^- := \{x \in \mathcal{P}([d] \setminus \{i\}) : x \in S\}$.

We may express the i -sections of $C_i(S)$ and $D_i(S)$ purely in terms of the i -sections of S . Indeed, $(C_i(S))_i^- = S_i^-$ and $(C_i(S))_i^+ = S_i^- \cap S_i^+$. Similarly, we have that $(D_i(S))_i^- = S_i^- \cup S_i^+$ and $(D_i(S))_i^+ = S_i^+$.

We also define, for $S \subseteq \mathcal{P}[d]$ and $i \in [d]$, four related subsets $T, U, V, W \subseteq \mathcal{P}([d] \setminus \{i\})$:

$$\begin{aligned} T &= T_{S,i} = S_i^+ \cap S_i^- = \{x \in \mathcal{P}([d] \setminus \{i\}) : x \in S \text{ and } x \cup i \in S\}, \\ U &= U_{S,i} = S_i^- \setminus S_i^+ = \{x \in \mathcal{P}([d] \setminus \{i\}) : x \in S \text{ and } x \cup i \notin S\}, \\ V &= V_{S,i} = S_i^+ \setminus S_i^- = \{x \in \mathcal{P}([d] \setminus \{i\}) : x \notin S \text{ and } x \cup i \in S\}, \\ W &= W_{S,i} = \{x \in \mathcal{P}([d] \setminus \{i\}) : x \notin S, x \cup i \notin S\}. \end{aligned}$$

Given a subset A of $\mathcal{P}([d] \setminus \{i\})$, we write $A \times \{i\}$ for the set $\{a \cup \{i\} : a \in A\}$.

These sets give us another way of viewing C_i and D_i . Indeed, as shown in Figure 3.1, $S = (T \cup U) \cup ((T \cup V) \times \{i\})$. Similarly, $C_i(S) = S \setminus (V \times \{i\})$ and $D_i(S) = S \cup U$.

3.3.2 Edge Version of Main Result

For sets, $S_1, S_2 \subseteq V(Q_d)$, we denote by $\partial_e(S_1, S_2)$ the set of edges with one endpoint in S_1 and one endpoint in S_2 . Similarly, we write $\vec{\partial}_e(S_1, S_2)$ for the set of edges with the smaller endpoint in S_1 and the larger endpoint in S_2 , i.e. edges \vec{uv} with $u \in S_1$ and $v \in S_2$. One can see that $\partial_e(S) = \partial_e(S, S^c)$ and $\vec{\partial}_e(S) = \vec{\partial}_e(S, S^c)$.

We now proceed in proving Theorem 3.5, which, as noted in the Introduction, implies Theorem 3.4 by Lemma 3.2.

Proof of Theorem 3.5. The majority of the proof is contained in the following key lemma.

Lemma 3.19. *For any set S , and all i , $|\vec{\partial}_e(S)| \geq \frac{1}{2}|\vec{\partial}_e(C_i(S))| + \frac{1}{2}|\vec{\partial}_e(D_i(S))|$.*

Proof of Lemma 3.19. For convenience, in the proof of this lemma, we write C for $C_i(S)$ and D for $D_i(S)$. It is easy to see that the contribution to $\vec{\partial}_e(S)$ from edges along the i direction, is exactly the same as the contribution to $\vec{\partial}_e(C)$ and to $\vec{\partial}_e(D)$.

Firstly, we can see that since D is a superset of S , any element of the directed edge boundary of S is in the directed edge boundary of D unless its larger endpoint is in $D \setminus S$. Thus $\vec{\partial}_e(S) \setminus \vec{\partial}_e(D) = \vec{\partial}_e(T \cup U, V)$.

Conversely, an element of $\vec{\partial}_e(D)$ is an element of $\vec{\partial}_e(S)$ unless its smaller endpoint is in $D \setminus S$. Therefore, $\vec{\partial}_e(D) \setminus \vec{\partial}_e(S) = \vec{\partial}_e(V, W)$.

Similar arguments show that $\vec{\partial}_e(S) \setminus \vec{\partial}_e(C) = \vec{\partial}_e(V \times \{i\}, (W \cup U) \times \{i\})$ and that $\vec{\partial}_e(C) \setminus \vec{\partial}_e(S) = \vec{\partial}_e(T \times \{i\}, V \times \{i\})$.

Thus

$$\begin{aligned} |\vec{\partial}_e(S)| - |\vec{\partial}_e(D)| &= |\vec{\partial}_e(T \cup U, V)| - |\vec{\partial}_e(V, W)| \\ &\geq |\vec{\partial}_e(T, V)| - |\vec{\partial}_e(V, W \cup U)| \\ &= |\vec{\partial}_e(C)| - |\vec{\partial}_e(S)|. \end{aligned}$$

Therefore, $|\vec{\partial}_e(S)| \geq \frac{1}{2} (|\vec{\partial}_e(C)| + |\vec{\partial}_e(D)|)$.

□

Given S such that $A \subseteq S \subseteq B^c$, for $i = 1, \dots, d$, we successively apply either C_i or D_i . By Lemma 3.19, we may make the choice of C_i or D_i to end with a set S' with $|\vec{\partial}_e(S')| \leq |\vec{\partial}_e(S)|$. By Lemma 3.18, S' must be i -down for all i and thus is a down set. By Observation 3.17, we have that the set satisfies the required containments. Since, the directed edge boundary of a down set is the same as the edge boundary, Theorem 3.21 finishes the proof. □

We note that we do not in fact require the full strength of Lemma 3.19; that $|\vec{\partial}_e(S)|$ is greater than the minimum of $|\vec{\partial}_e(C_i(S))|$ and $|\vec{\partial}_e(D_i(S))|$ suffices.

3.3.3 Vertex Version of Main Result

The proof of Theorem 3.11 is very similar to the proof of the edge version, but with a slightly different calculation.

Proof of Theorem 3.11. Again, the bulk of the proof is in the following lemma.

Lemma 3.20. *For any set S , and all i , $|\vec{\partial}_v(S)| \geq \frac{1}{2} \left(|\vec{\partial}_v(C_i(S))| + |\vec{\partial}_v(D_i(S))| \right)$.*

Proof. Once more, we write C for $C_i(S)$ and D for $D_i(S)$. Additionally, we write $h(S) = S \cup \vec{\partial}_v(S)$.

Since C is a subset of S , any vertex in the directed vertex boundary of C is in the directed vertex boundary of S unless it is in $S \setminus V$. Thus $\vec{\partial}_v(C) \setminus \vec{\partial}_v(S) = \left(\vec{\partial}_v(T) \cap V \right) \times \{i\}$.

On the other hand, any vertex in $\vec{\partial}_v(S)$ but not in $\vec{\partial}_v(C)$ must neighbour a vertex in $S \setminus C$ and thus $\vec{\partial}_v(S) \setminus \vec{\partial}_v(C) = \vec{\partial}_v(V) \times \{i\} \setminus \vec{\partial}_v(C)$. Since the set of vertices in $\vec{\partial}_v(C)$ that contain i is $(U \times \{i\}) \cup h(T \times \{i\})$, we may conclude that $\vec{\partial}_v(S) \setminus \vec{\partial}_v(C) = \left(\vec{\partial}_v(V) \setminus (h(T) \cup U) \right) \times \{i\}$.

Similarly, we have that $\vec{\partial}_v(S) \setminus \vec{\partial}_v(D) = \vec{\partial}_v(T \cup U) \cap V$ and that $\vec{\partial}_v(D) \setminus \vec{\partial}_v(S) = \vec{\partial}_v(V) \setminus h(T \cup U)$.

Since V is disjoint from $T \cup U$, we see that $\vec{\partial}_v(T) \cap V \subseteq \vec{\partial}_v(T \cup U) \cap V$. Thus

$$\begin{aligned} |\vec{\partial}_v(S)| - |\vec{\partial}_v(D)| &= |\vec{\partial}_v(T \cup U) \cap V| - |\vec{\partial}_v(V) \setminus h(T \cup U)| \\ &\geq |\vec{\partial}_v(T) \cap V| - |\vec{\partial}_v(V) \setminus (h(T) \cup U)| \\ &= |\vec{\partial}_v(C)| - |\vec{\partial}_v(S)|. \end{aligned}$$

Therefore, $|\vec{\partial}_v(S)| \geq \frac{1}{2} \left(|\vec{\partial}_v(C)| + |\vec{\partial}_v(D)| \right)$.

□

Given S , we can use Lemma 3.20, for $i = 1, \dots, d$, and successively apply either D_i or C_i to yield a set S' with $|\vec{\partial}_v(S')| \leq |\vec{\partial}_v(S)|$. By Lemma 3.18, S' must be i -down for all i and thus is a down set. By Observation 3.17, we have that S' satisfies the required containments.

□

We now deduce Theorem 3.10, on vertex-disjoint directed paths.

Proof of Theorem 3.10. Let $F = \{xy \in E(Q_d) : x \in A, y \in B\}$. We apply the directed version of Menger's Theorem to the directed graph $G = \overrightarrow{Q_d} - F$. It tells us that the number of paths in G from A to B , with vertex-disjoint interiors is the same as the minimum vertex cut separating A from B in G . This is the same as $\min \left\{ |\overrightarrow{\partial}_v(S)| : A \subseteq S \subseteq B^c \right\} - |\{x \in B : d(x, y) = 1, \text{ for some } y \in A\}|$. Theorem 3.11 implies this is the same as $\min \left\{ |\partial_v(S)| : A \subseteq S \subseteq B^c \right\} - |\{x \in B : d(x, y) = 1, \text{ for some } y \in A\}|$, since the directed boundary is minimized by a down set. Observation 3.8 concludes the proof. \square

3.4 Bollobás and Leader's Theorems

For completeness, in this section we discuss the full versions of Bollobás and Leader's theorems on edge-disjoint and vertex-disjoint paths, and correct the minor mistakes we found in that paper.

3.4.1 Undirected Edge-Disjoint Paths

In this subsection, we state the amended form of Bollobás and Leader's full version of Theorem 3.3, and give its proof, as well as that of Lemma 3.2, upon which it relies.

First we give an approximation to $|\partial_e(I_t)|$, the size of the edge boundary of an initial segment of binary order. Chung, Füredi, Graham and Seymour [26] observed the following good lower bound:

$$e(x) = e_d(x) = \begin{cases} x(d - \log_2 x) & \text{if } x \leq 2^{d-1}, \\ (2^d - x)(d - \log_2(2^d - x)) & \text{if } x > 2^{d-1}. \end{cases}$$

This function $e(x)$ is easier to work with than $|\partial_e(I)|$, as there is a greater degree of monotonicity, and it plays a key role in the proof of the following theorem. Note that $e(2^k) = (d - k)2^k$, and hence the below theorem is a generalisation of Theorem 3.3.

Theorem 3.21 (Bollobás-Leader [17]). *Let A and B be disjoint non-empty subsets of $V(Q_d)$. Then there is a family of at least $\min \{e(|A|), e(|B|), 2^{d-1}\}$ edge-disjoint directed paths from A to B .*

The function e is monotone increasing up to $x = 2^d/\exp(1)$, it then decreases until $x = 2^{d-1}$, and is symmetric about this point. Although the argument of [17] is essentially correct, it was incorrectly stated that e is increasing up to $x = 2^{d-1}$, leading to the erroneous bound $\min\{e(|A|), e(|B|)\}$ in the theorem. Note that this only differs from the bound above when $|A|$ and $|B|$ are both larger than 2^{d-2} .

We may now define concretely the function $BL_e(x, y)$ that we left undefined in Section 3.1.1: $BL_e(|A|, |B|) = \min\{e(|A|), e(|B|), 2^{d-1}\}$.

We give the proof of Lemma 3.2 and then deduce Theorem 3.21. Note that only the first part of the Lemma is required for Theorem 3.21, but the second part was required for our Theorem 3.4, so we prove this part in the greater detail.

Proof of Lemma 3.2. To prove the second part, we apply Theorem 3.14 to \vec{Q}_d . Given an edge cut, write S for the component containing A in the graph formed by deleting the edge cut. Clearly $A \subseteq S \subseteq B^c$. If the cut is minimal, then $\vec{\partial}_e(S)$ is the whole cut and its size is precisely $|\vec{\partial}_e(S)|$.

Similarly, to prove the first part, we apply Theorem 3.14 to the graph Q_d . Again, a minimal edge cut corresponds to $\partial_e(S)$ for some S satisfying $A \subseteq S \subseteq B^c$, which concludes the proof. \square

Proof of Theorem 3.21. By Lemma 3.2, we may choose S with $p_e(A) = \partial_e(S)$ and $A \subseteq S \subseteq B^c$. Recall that $|\partial_e(S)| \geq e(|S|)$. If $e(|S|) \geq 2^{d-1}$, we are done. If not, since $e(2^{d-2}) = e(2^{d-1}) = e(3 \cdot 2^{d-2}) = 2^{d-1}$, we have that $|A| \leq |S| < 2^{d-2}$ or $|B^c| \leq |S^c| < 2^{d-2}$. In either case, monotonicity of e in these intervals completes the proof. \square

3.4.2 Matchings and Undirected Vertex-Disjoint Paths

In this section we give the full version of Bollobás and Leader's lower bound on $p_v(Q_d, A, B)$, for all values of $|A|$ and $|B|$, as well as an amended proof. We also give an amended proof of another result in [17] on the size of the maximum matching between a set of vertices of Q_d and its complement.

We first follow [17] and define the function $b(x)$, used as a lower bound for $|\partial_v(S)|$, where S is a set of size x . For all d , we may write any smaller positive number x uniquely in the

form $\sum_{i=0}^k \binom{d}{k} + \alpha \binom{d}{k+1}$, for some $0 \leq k \leq d$ and $0 \leq \alpha < 1$. We then define

$$b(x) = b_d(x) = (1 - \alpha) \binom{d}{k+1} + \alpha \binom{d}{k+2}.$$

This allows us to state the full vertex-disjoint paths result from [17].

Theorem 3.22 (Bollobás-Leader [17]). *Let A and B be disjoint non-empty subsets of Q_d . Then there is a family of at least the minimum of $b(|A|)$ and $b(|B|)$ paths from A to B with vertex-disjoint interiors.*

Note that $b\left(\sum_{i=0}^k \binom{d}{k}\right) = \binom{d}{k+1}$, so this theorem agrees with the special case, Theorem 3.9, stated above. We may also now define the function BL_v , left undefined in Section 3.1.2. Indeed, $BL_v(x, y) = \min\{b(x), b(y)\}$.

In the case where d is even and $|A|$ and $|B|$ are very close to 2^{d-1} , the proof in [17] contains a small error in a calculation, although the theorem is correct as stated. For completeness, we give the full, amended proof here, despite the change being a minor one. Indeed, the change is simply using a slightly stronger lower bound for $\vec{\partial}_v$ than b , on one occasion.

The function $b(x)$ is increasing up to $x = \sum_{i=1}^{\lceil d/2 \rceil} \binom{d}{i}$, and is decreasing thereafter. If d is odd, this is equal to 2^{d-1} . If d is even, however, this is slightly less than 2^{d-1} . It was incorrectly stated in [17] that b is increasing up to 2^{d-1} in both cases.

For the amended proof, we will require a weak version of the Kruskal–Katona Theorem, due to Lovász [68]. Here, and in what follows, we write $\overleftarrow{\partial}_v A := \{x \setminus \{i\} : i \in x \text{ and } x \in A\}$. This is sometimes known as the *lower shadow* of A .

Theorem 3.23 (Lovász [68]). *Let $A \subseteq [d]^{(r)}$. Write $|A| = \binom{x}{r}$, $x \in \mathbb{R}, x > r - 1$. Then $|\overleftarrow{\partial}_v A| \geq \binom{x}{r-1}$.*

The proof of Theorem 3.22 is somewhat analogous to that of Lemma 3.2, but is complicated slightly by the fact there may be some edges from A to B , so we cannot directly apply flow theorems in Q_d .

Proof of Theorem 3.22. We let $F = \{xy \in E(Q_d) : x \in A, y \in B\}$ and we will apply Theorem 3.15, i.e Vertex Menger’s Theorem, in the graph $G = Q_d - F$. Writing A_1 for

$\{x \in A : xy \in F \text{ for some } y \in B\}$, and similarly B_1 for $\{x \in B : xy \in F \text{ for some } y \in A\}$, it is clear that $|F| \geq \max(|A_1|, |B_1|)$. It is therefore sufficient to show that any set $C \subseteq Q_d$ separating A from B in G has size at least $\min\{b(|A|), b(|B|)\} - \max\{|A_1|, |B_1|\}$.

Let S be a subset of $V(Q_d)$ that separates A from B in G . Let A' be the union of the components that contain vertices of A in the graph $G - S$, and define B' to be the union of all other components, i.e. $B = V(Q_d) \setminus (A' \cup S)$. We may assume that $|A'| \leq |B'|$. Since A' and B' are disjoint, we get that $|A'| \leq 2^{d-1}$. If d is even and $|A'| \leq \sum_{i=0}^{d/2-1} \binom{d}{i}$ or if d is odd, then $\partial_v(A') \geq b(|A|)$, by the monotonicity of b up to this point. Since also $\partial_v(A') \subseteq S \cup B_1$, we are done in this case.

In the other case, d is even, $|A'| = \sum_{i=0}^{d/2-1} \binom{d}{i} + \alpha \binom{d}{d/2}$ and $|B'| = \sum_{i=0}^{d/2-1} \binom{d}{i} + \beta \binom{d}{d/2}$, for some $0 < \alpha, \beta < 1$. Since A' and B' are disjoint, $\alpha \leq 1/2$.

Recall that $|\partial_v A'| \geq |\partial_v I|$, where I is an initial segment of the simplicial order, with $|I| = \sum_{i=0}^{d/2-1} \binom{d}{i} + \alpha \binom{d}{d/2}$. We write I_0 for $|I \cap [d]^{(d/2)}|$. It is easy to see that $|I_0| = \alpha \binom{d}{d/2}$ and that $|\vec{\partial}_v(I)| = (1-\alpha) \binom{d}{d/2} + |\vec{\partial}_v(I_0)|$. We will show that $|\vec{\partial}_v I_0| \geq |I_0|$. Let $J_0 = \{x^c : x \in I_0\}$, a subset of $[d]^{(d/2)}$. Note that $|\overleftarrow{\partial}_v(J_0)| = |\vec{\partial}_v I_0|$. Choose x such that $|J_0| = \binom{x}{d/2}$. Since $\binom{d-1}{d/2} = \frac{1}{2} \binom{d}{d/2} \geq |J_0|$, we have that $x \leq d-1$.

Thus, by Theorem 3.23 we have:

$$\begin{aligned} |\overleftarrow{\partial}_v(J_0)| - |J_0| &\geq \binom{x}{d/2-1} - \binom{x}{d/2} \\ &\quad \frac{x(x-1) \cdots (x-d/2+2)}{(d/2)!} [d/2 - (x-d/2+1)] \\ &\geq 0. \end{aligned}$$

This implies that $|\vec{\partial}_v(I_0)| \geq |I_0|$, and so $|\partial A'| \geq \binom{d}{d/2}$. Since $B_1 = B' \cap \partial A'$, we have that $|B_1| \geq |A'| + |\partial A'| + |B'| - 2^d \geq (\alpha + \beta) \binom{d}{d/2}$. Since $|S| = (1 - \alpha - \beta) \binom{d}{d/2}$ and $b(|A|) \leq \binom{d}{d/2}$, we are done. □

Note that the proof implies Observation 3.8 with no extra work.

Essentially the same issue occurs in the proof of another theorem from [17]. For completeness, we state the theorem below and give an amended proof.

The *surface* of $S \subseteq \mathcal{P}[d]$, written $\sigma(S)$, is the set of vertices of S adjacent to a vertex

in S^c . In other words, $\sigma(S) = \{x \in S : \exists y \in S^c, d(x, y) = 1\}$. The reader may notice a similarity to the definition of $\partial_v(S)$. Indeed, $\sigma(S) = \partial_v(S^c)$.

We write $s(x) = s_d(x) = (1 - \alpha)\binom{d}{k} + \alpha\binom{d}{k+1}$, where $x = \sum_{i=0}^k \binom{d}{i} + \alpha\binom{d}{k+1}$ for some $\alpha \in [0, 1)$. The relationship between σ and ∂_v implies that $\sigma(A) \geq s(|A|)$, for all sets A .

Bollobás and Leader showed the following, essentially best possible, bound on the size of matchings between two complementary sets, in terms of the size of the smaller set.

Theorem 3.24 (Bollobás-Leader [17]). *Let A be a subset of Q_d with $|A| \leq 2^{d-1}$. Then there is a matching from A to A^c of size at least $s(|A|)$.*

Again, the theorem was stated correctly, but in [17] it was incorrectly claimed that $b(x)$ is increasing up to $x = 2^{d-1}$. Once more, the fix is a small part of the proof; the remainder comes directly from [17].

Proof of Theorem 3.24. By Defect Hall's Theorem, there is a matching of size $s(|A|)$ if there is no $B \subset A$ with $|\partial B \cap A^c| < |B| - (|A| - s(|A|))$. Suppose such a B existed. Then we must have that: $|B| \geq |A| - s(|A|)$.

If $|B| \leq \sum_{i=0}^{\lfloor d/2 \rfloor} \binom{d}{i}$, then by monotonicity of b up to this point, $b(|B|) \geq b(|A| - s(|A|))$. Note that $b(|A| - s(|A|)) = s(|A|)$, by definition of b and s , so $b(|B|) \geq s(|A|)$.

Otherwise, the assumption on the size of $|A|$ implies that d is even and that $|A| = \sum_{i=0}^{d/2} \binom{d}{i} + \alpha\binom{d}{d/2}$ and $|B| = \sum_{i=0}^{d/2} \binom{d}{i} + \beta\binom{d}{d/2}$, for some $0 < \beta \leq \alpha \leq 1/2$. Thus,

$$b(|B|) = (1 - \beta)\binom{d}{d/2} + \beta\binom{d}{d/2+1}, \quad s(|A|) = (1 - \alpha)\binom{d}{d/2-1} + \alpha\binom{d}{d/2}.$$

Therefore

$$b(|B|) - s(|A|) = (1 - \beta - \alpha)\binom{d}{d/2} - (1 - \beta - \alpha)\binom{d}{d/2+1}.$$

Hence in this case, we also have $b(|B|) \geq s(|A|)$. This implies that $|\partial_v B \cap A^c| \geq s(|A|) - |A \setminus B|$, concluding the proof. \square

We note that the only amendment in the proof was in dealing separately with sets B of size greater than $\sum_{i=0}^{\lfloor d/2 \rfloor} \binom{d}{i}$.

3.5 Generalisation to Grids

In this subsection, we prove Theorems 3.6 and 3.12, that is, we generalise to grids our theorems on edge-disjoint and vertex-disjoint directed paths in the hypercube, again proceeding via isoperimetric-type inequalities. Indeed, as remarked in the introduction, the approach is similar, so we skip most of the details, highlighting only the changes. In particular, we only comment on the proof of the isoperimetric-type statements, since the deduction of Theorems 3.6 and 3.12 is essentially identical to the deduction of Theorems 3.4 and 3.10 from Theorems 3.5 and 3.11 respectively.

We will find it useful to view P_m^d as m copies of P_m^{d-1} . In particular, given an $S \subseteq P_m^d$, we write S_i^k for the set of tuples in S whose i^{th} coordinate is k . In other words, $S_i^k = \{(x_1, x_2, \dots, x_d) \in S : x_i = k\}$. We call the S_i^k the i -sections of S .

For $S \subseteq [m]^d$, $i \in [d]$ and $t \in [m-1]$, we define two compressions, C_i^t and D_i^t , each of which preserve all of the i -sections of S , except for S_i^{t+1} and S_i^t respectively. More specifically,

$$(C_i^t(S))_i^k = \begin{cases} S_i^t \cap S_i^{t+1} & \text{if } k = t+1 \\ S_i^k & \text{if } k \neq t+1 \end{cases},$$

and

$$(D_i^t(S))_i^k = \begin{cases} S_i^t \cup S_i^{t+1} & \text{if } k = t \\ S_i^k & \text{if } k \neq t \end{cases}.$$

Note that when $m = 2$, the grid P_m^d is isomorphic to Q_d and these compressions correspond to the compressions in Section 3.3.

It is easy to see that the natural extension of Observation 3.17 holds for these grid compressions.

Observation 3.25. *If A is a down set, B is an up set and $A \subseteq S \subseteq B^c$, then, for $i \in [d]$, we have that $A \subseteq C_i(S) \subseteq S \subseteq D_i(S) \subseteq B^c$.*

In the hypercube case, we were able to deduce that all sequences of compressions converge by showing C_i and D_i produce sets that are i -down. This method does not generalize, so we use a different approach.

For ease of notation, we will appeal to symmetry and focus mainly on C_d and D_d .

Given a set $A \subseteq [m]^{d-1}$, we write $A \times \{k\}$ for the set $\{(a_1, a_2, \dots, a_{d-1}, k) \mid (a_1, \dots, a_{d-1}) \in A\}$.

Similar to our definitions in Section 3.3, for $S \subseteq [m]^d$ and $t \leq m-1$, we define four related subsets $T, U, V, W \subseteq [m]^{d-1}$:

$$\begin{aligned}
T &:= S_d^t \cap S_d^{t+1} \\
&= \{(x_1, \dots, x_{d-1}) \in [m]^{d-1} : (x_1, \dots, x_{d-1}, t) \text{ and } (x_1, \dots, x_{d-1}, t+1) \in S\}, \\
U &:= S_d^t \setminus S_d^{t+1} \\
&= \{(x_1, \dots, x_{d-1}) \in [m]^{d-1} : (x_1, \dots, x_{d-1}, t) \in S \text{ and } (x_1, \dots, x_{d-1}, t+1) \notin S\}, \\
V &:= S_d^t \setminus S_d^{t+1} \\
&= \{(x_1, \dots, x_{d-1}) \in [m]^{d-1} : (x_1, \dots, x_{d-1}, t) \notin S \text{ and } (x_1, \dots, x_{d-1}, t+1) \in S\}, \\
W &:= \{(x_1, \dots, x_{d-1}) \in [m]^{d-1} : (x_1, \dots, x_{d-1}, t) \notin S, (x_1, \dots, x_{d-1}, t+1) \notin S\}.
\end{aligned}$$

Notice that the definitions here are extensions of the $i = d$ case of the definitions in Section 3.3.

3.5.1 Edge-Disjoint Paths in a Grid

Theorem 3.6 follows immediately from the below analogue of Theorem 3.5. We do not show the details of this as it is identical to the deduction of Theorem 3.4 from Theorem 3.5.

Theorem 3.26. *Suppose A and B are disjoint subsets of $[m]^d$, where A is a down set, and B is an up set. Then $\min \left\{ |\vec{\partial}_e(S)| : A \subseteq S \subseteq B^c \right\}$ is attained by a down set. Thus $\min \left\{ |\vec{\partial}_e(S)| : A \subseteq S \subseteq B^c \right\} = \min \{ |\partial_e(S)| : A \subseteq S \subseteq B^c \}$.*

We will use the following lemma, which corresponds to Lemma 3.19. Note that unlike the hypercube case, we do require the full strength of it. The main difference between the proof of this lemma and that of Lemma 3.19 is that C_i^t and D_i^t may now alter the contribution to $\vec{\partial}_e(S)$ in the i^{th} direction; we show that it may only reduce this contribution.

Lemma 3.27. *If $S \subseteq [m]^d$, then for all $i \in [d]$ and $t \in [m-1]$, we have that $|\vec{\partial}_e(S)| \geq \frac{1}{2} |\vec{\partial}_e(C_i^t(S))| + \frac{1}{2} |\vec{\partial}_e(D_i^t(S))|$.*

Proof of Lemma 3.27. Without loss of generality, we may assume $i = d$. We write C for $C_d^t(S)$ and D for $D_d^t(S)$. For ease of notation, for $A, B \subseteq [m]^d$, we write $\vec{\partial}_e(A, B)$ for the set of edges whose smaller endpoint is in A and whose larger endpoint is in B .

Since D is a superset of S , any element of the directed edge boundary of S is contained in the directed edge boundary of D unless its larger endpoint is in $D \setminus S$. Thus $\vec{\partial}_e(S) \setminus \vec{\partial}_e(D) \geq \vec{\partial}_e((T \cup U) \times \{t\}, V \times \{t\})$; equality holds only if there are no edges between $S_d^{t-1} \times \{t-1\}$ and $V \times \{t\}$.

Conversely, an element of $\vec{\partial}_e(D)$ is an element of $\vec{\partial}_e(S)$ unless its smaller endpoint is in $D \setminus S$. Hence we have $\vec{\partial}_e(D) \setminus \vec{\partial}_e(S) = \vec{\partial}_e(V \times \{t\}, W \times \{t\})$.

Similar arguments show that $\vec{\partial}_e(S) \setminus \vec{\partial}_e(C) \geq \vec{\partial}_e(V \times \{t+1\}, (W \cup U) \times \{t+1\})$ and $\vec{\partial}_e(C) \setminus \vec{\partial}_e(S) = \vec{\partial}_e(T \times \{t+1\}, V \times \{t+1\})$.

Thus

$$\begin{aligned} |\vec{\partial}_e(S)| - |\vec{\partial}_e(D)| &\geq |\vec{\partial}_e(T \cup U, V)| - |\vec{\partial}_e(V, W)| \\ &\geq |\vec{\partial}_e(T, V)| - |\vec{\partial}_e(V, W \cup U)| \\ &\geq |\vec{\partial}_e(C)| - |\vec{\partial}_e(S)|. \end{aligned}$$

Therefore, $|\vec{\partial}_e(S)| \geq \frac{1}{2} (|\vec{\partial}_e(C)| + |\vec{\partial}_e(D)|)$.

□

From this, we now deduce Theorem 3.26.

Proof of Theorem 3.26. Fix an S with $A \subseteq S \subseteq B^c$. For fixed i, t , either one of C_i^t and D_i^t reduces the size of the edge boundary or both of them keep it the same, by the lemma. Now, unless both compressions preserve S , we have that $|C_i^t(S)| < |S|$. Thus we may transform S into a set stable under all compressions, i.e. a down set, by applying only compressions that reduce $(|\vec{\partial}_e(S)|, |S|)$, in lexicographic order (i.e. $(x_1, x_2) < (y_1, y_2)$ if $x_1 < y_1$ or $x_1 = y_1$ and $x_2 < y_2$). Thus there is some down set S with $|\vec{\partial}_e(S')| \leq |\vec{\partial}_e(S)|$, concluding the proof. □

3.5.2 Vertex-Disjoint Paths in a Grid

Again we proceed via an isoperimetric-type statement, although we will not show how it implies Theorem 3.12.

Theorem 3.28. *Let A be a non-empty down set and B be a non-empty up set, both non-empty subsets of P_m^d . Suppose $A \subseteq S \subseteq B^c$, then there exists a down set S' satisfying $A \subseteq S' \subseteq B^c$ with $\vec{\partial}_v(S) \geq \vec{\partial}_v(S')$.*

We skip the proof of this theorem, noting only that it follows from the below lemma in exactly the same way that Lemma 3.27 follows from Theorem 3.26.

Lemma 3.29. *If $S \subseteq [m]^d$, then for all $i \in [d]$ and $j \in [m-1]$, we have that $|\vec{\partial}_v(S)| \geq \frac{1}{2}|\vec{\partial}_v(C_i^j(S))| + \frac{1}{2}|\vec{\partial}_v(D_i^j(S))|$.*

Proof. Once again, we assume $i = d$, without loss of generality. We write C for $C_d^t(S)$ and D for $D_d^t(S)$. Additionally, we write $h(S) = S \cup \vec{\partial}_v(S)$.

Since C is a subset of S , any vertex in $\vec{\partial}_v(C)$ is in $\vec{\partial}_v(D)$ unless it lies in $S \setminus C$. Thus $\vec{\partial}_v(C) \setminus \vec{\partial}_v(S) = (\vec{\partial}_v(T) \cap V) \times \{t+1\}$.

Similarly, any vertex in $\vec{\partial}_v(S) \setminus \vec{\partial}_v(C)$ must neighbour a vertex in $S \setminus C$. Thus $\vec{\partial}_v(S) \setminus \vec{\partial}_v(C) \geq (\vec{\partial}_v(V) \setminus (h(T) \cup U)) \times \{t+1\}$, where equality holds only if $V \subseteq S^{t+2}$.

Likewise, we have that $\vec{\partial}_v(S) \setminus \vec{\partial}_v(D) \geq (\vec{\partial}_v(T \cup U) \cap V) \times \{t\}$ and $\vec{\partial}_v(D) \setminus \vec{\partial}_v(S) = (\vec{\partial}_v(V) \setminus (h(T) \cup U)) \times \{t\}$.

Note that V is disjoint from $T \cup U$ and so $\vec{\partial}_v(T) \cap V \subseteq \vec{\partial}_v(T \cup U) \cap V$.

Simple rearrangement shows that $|\vec{\partial}_v(S)| - |\vec{\partial}_v(D)| \geq |\vec{\partial}_v(C)| - |\vec{\partial}_v(S)|$; this completes the proof of the lemma. \square

Chapter 4

Local Biclique Decompositions

4.1 Introduction

4.1.1 Biclique covers and partitions

Given a family of graphs, \mathcal{F} , an \mathcal{F} -cover of a graph G is a collection of subgraphs of G , each isomorphic to an element of \mathcal{F} , such that every edge of G is in at least one of the subgraphs. If additionally each edge of G is only covered once, we call the collection of bicliques an \mathcal{F} -partition of G . Several problems related to minimal \mathcal{F} -covers and minimal \mathcal{F} -partitions have been studied. See for instance [21] for the case where \mathcal{F} is the set of cliques, and [32] for $\mathcal{F} = \{K_2, K_3\}$.

In this chapter, we discuss the case where \mathcal{F} is the family of complete bipartite graphs, which are known as *bicliques*. A biclique with parts of size n and m is written $K_{n,m}$. Since bicliques are determined by the bipartition, we may write them as (α, β) , where α and β are the vertex classes in the bipartition. When \mathcal{F} is the collection of bicliques, \mathcal{F} -covers and \mathcal{F} -partitions are called biclique covers and biclique partitions respectively. A biclique cover consisting of k bicliques is known as a k -cover; similarly, k -partitions are biclique partitions consisting of k bicliques.

The *biclique cover number* of G , $\text{bc}(G)$, is the least k for which there exists a k -cover of G . Analogously, the *biclique partition number* of G , $\text{bp}(G)$, is the least k for which there exists a k -partition of G . The biclique cover number and biclique partition number have been studied extensively, see for instance [77] and [53], motivated by applications to many other areas such as combinatorial geometry [1], communication complexity [65], network addressing [39] and even immunology [78].

One of the early results on the biclique partition number is the Graham–Pollak Theorem [39]—see [89] for an elegant proof—which states that $\text{bp}(K_n) = n - 1$. In contrast to this, it is easy to show that $\text{bc}(K_n) = \lceil \log n \rceil$. (Here and in the remainder of this chapter, \log refers to \log_2). These results demonstrate that the trivial inequality $\text{bc}(G) \leq \text{bp}(G)$ may be quite loose: $\text{bp}(K_n) \approx 2^{\text{bc}(K_n)}$.

In Section 4.3, we investigate how large bp can be in terms of bc . Indeed, we prove that if $\text{bc}(G) = t$, then $\text{bp}(G) \leq (3^t - 1)/2$ and exhibit a graph showing that this is tight.

We call a biclique cover (resp. partition) *r-local* if every vertex is in at most r of the

bicliques involved in the cover (resp. partition) of G . The *local biclique cover number*, $\text{lbc}(G)$, is the least r such that there exists an r -local cover of the graph. Similarly, the *local biclique partition number*, $\text{lbp}(G)$, is the least r such that there exists an r -local partition of the graph.

A closely related concept is that of *weight* of a biclique cover, which is the sum of the orders of the bicliques in a cover. The *weight biclique cover number*, written $\text{wbc}(G)$, is the minimum weight of a biclique cover of G ; similarly, the *weight biclique partition number* is the minimum weight of a biclique partition of G and is written $\text{wbp}(G)$. It is easy to see that $\text{lbc}(G) \geq \frac{\text{wbc}(G)}{|G|}$.

Variants of lbp and lbc , such as wbc and wbp , have long been studied, starting in 1967 with Katona and Szemerédi [57] in connection with diameter 2 graphs with few edges. An easy corollary of their result is:

$$\text{lbc}(K_n) = \text{lbp}(K_n) = \lceil \log n \rceil,$$

which was also shown explicitly by Dong and Liu [29]. This result is perhaps surprising, given that $\text{bc}(K_n)$ and $\text{bp}(K_n)$ differ so vastly.

The extremal order of these numbers has been known for a long time. Indeed, there is a constant c_1 such that $\text{lbp}(G) \leq \frac{c_1 n}{\log n}$, for all graphs G on n vertices, while there is some constant c_2 such that for all n , there exists a graph G on n vertices with $\text{lbc}(G) \geq \frac{c_2 n}{\log n}$. These bounds were shown implicitly by Lupanov [69] in 1956 (see Lemma 1.2 in [52] for a more accessible version) and shown more explicitly by Erdős and Pyber [34] and Csirmaz, Ligeti and Tardos [27], the latter result motivated by applications to secret sharing schemes. These results may be interpreted as saying that if $\text{lbc}(G)$ is large compared to $|G|$, then $\text{lbp}(G)$ is not much larger.

This is one motivation for seeking to bound lbp in terms of lbc , hoping for a similar result to our previously mentioned bound for bc in terms of bp . Our main result of this chapter, however, shows that no such bound is possible. More specifically, in Section 4.4 we exhibit, for all $k \geq 2$, a graph G with $\text{lbc}(G) = 2$, but $\text{lbp}(G) \geq k$. Since a 1-local cover is itself a partition, this is the strongest possible result along these lines.

Given this, it is natural to ask ‘how fast’ can $\text{lbp}(G)$ tend to infinity if $\text{lbc}(G) = 2$. One way of formalizing this question is to ask ‘if G has a 2-local m -cover, how large can $\text{lbp}(G)$

be?’. We answer this question, up to a constant factor, in Section 4.5. Although this chapter is based on work of the author in [81], the bound given here improves that in the paper by a factor of 2.

We leave open the equally natural question of ‘if $\text{lbc}(G) = 2$ and $\text{bc}(G) = m$, how large can lbp be?’. A second question that we leave open is whether all graphs have a single cover that is close to optimal for both bc and lbc . We discuss these questions in Section 4.6 and also note an area of computer science to which our result that lbc does not bound lbp , later labelled Theorem 4.7, has been applied, subsequent to its publication in [81].

4.1.2 Subcube Intersection Graph Representations

The biclique cover number discussed above has a natural interpretation in the setting of subcube intersection graphs.

An *intersection graph* of a family of subsets of some groundset has one vertex for each of the sets in the family, and an edge between two vertices if and only if the corresponding sets intersect. See [51] or [71] for more on intersection graphs for various set families. In this paper, we are interested in *subcube intersection graphs*, which are intersection graphs where all sets in the family are subcubes of a hypercube, Q_d . Note that two subcubes intersect if and only if they agree on all coordinates where both are fixed. See Johnson and Markström [48] for a more detailed background.

Let $I(n, d)$ be the set of all graphs on n vertices that are the intersection graph of some family of subcubes of a d -dimensional hypercube. It is easy to see that $G \in I(n, d)$ if and only if its complement has a d -cover, as was apparently first pointed out by Fishburn and Hammer [36], and also noted in [48]. Indeed, a representation of $G \in I(n, d)$ assigns a vector in $\{0, 1, *\}^d$ to each vertex. We generate a biclique for each of the d coordinates, having as one class all the vertices with a zero in this coordinate, and as the other class all vertices with a one in this coordinate. Then, the union of these bicliques has an edge exactly where G does not have an edge. This produces a d -cover of G^c and the converse is similar.

If G is a graph on n vertices, we write $\rho(G)$ for the smallest d such that $G \in I(n, d)$ —in other words, $\rho(G)$ is the smallest d such that G may be represented as an intersection graph of a family of subcubes of Q_d . By the previous paragraph, $\rho(G) = \text{bc}(G^c)$. We also write

$\rho(n) := \max\{\rho(G) : |G| = n\}$, where $|G|$ denotes the number of vertices of G . Clearly, this is also equal to $\max\{\text{bc}(G) : |G| = n\}$.

Similarly, we define $\tau(G)$ to be the smallest r such that G may be represented as an intersection graph of a family of r dimensional subcubes of some hypercube. We also let $\tau(n) = \max\{\tau(G) : |G| = n\}$. Using the relationship between subcube intersection graphs and biclique coverings, we may see that $\tau(G)$ is the least r such that G^c has a biclique cover where every vertex lies in all but r of the bicliques in the cover. (Note in this interpretation, bicliques with one empty class are allowed). This bears a resemblance to, but is distinct from $\text{lbc}(G^c)$ —the least r such that G^c has a cover with every vertex in no more than r bicliques of the cover.

Many authors have placed various bounds on ρ —see for instance [24], [30], [85] and [88]. The best known bounds are:

$$n - c \log n \leq \rho(n) \leq n - \lfloor \log n \rfloor + 1,$$

where c is some positive constant. The upper bound is due to Tuza [88] and the probabilistically proved lower bound to Rödl and Ruciński [85]. The question of obtaining similar bounds for $\tau(n)$ was posed by Johnson and Markström [48]. In Section 4.2, we shall prove a very close relationship between $\tau(n)$ and $\rho(n)$ and thus we obtain bounds on τ similar to those on ρ .

We remark briefly that the other biclique covering/partitioning measures may be interpreted in terms of subcube intersection graphs. For instance, $\text{lbc}(G^c)$ is the least r such that G can be represented as the intersection graph of a family of codimension r subcubes of some hypercube. Equally, $\text{bp}(G^c)$ is the least d such that G can be represented in $I(n, d)$ such that if u is not adjacent to v in G , then there is a unique i such that $\{u_i, v_i\} = \{0, 1\}$. We note also that the graphs we use for our bounds on bp and bc may be viewed naturally as subcube intersection graphs.

4.2 Relationship between τ and ρ

We shall prove the upper bounds and lower bounds on τ separately.

Lemma 4.1. *For any graph G , $\tau(G) \leq \rho(G)$, and hence $\tau(n) \leq \rho(n)$.*

Proof. Let G be a graph on n vertices with $\rho(G) = d$. This means that there are subcubes A_1, \dots, A_n of Q_d whose intersection graph is G . Let M (resp. m) be the maximum (resp. minimum) dimension of these subcubes. For all i , replace A_i by A'_i , a subcube of Q_{d+M-m} by appending 0's and *'s to the vector of A_i . We ensure that we add as many *'s as needed so the resultant vector has precisely M *'s and as many 0's as needed to give the vector length $(d + M - m)$. The intersection graph of the A'_i is G . This shows that G may be represented as an intersection graph of a family of subcubes of dimension M . Since $M \leq d$, the result follows. \square

Lemma 4.2. $\tau(n+1) \geq \rho(n)$.

Proof. Let G be a graph with vertices v_1, \dots, v_n and with $\rho(G) = \rho(n)$. Now form G' from G by adding a single vertex, v_{n+1} , adjacent to all vertices of G . Let A_1, \dots, A_n, A_{n+1} be r -dimensional subcubes of Q_d such that G' is the intersection graph of the A_i (the vertex v_i being represented by A_i) and such that r is equal to $\tau(G')$. Without loss of generality, let A_{n+1} be free in the first r coordinates, and hence fixed in the other $d-r$ coordinates. Where A_{n+1} is fixed, all the other A_i must be either free (have an asterisk in that coordinate) or have the same fixed value as A_{n+1} , as v_{n+1} is adjacent to all the other vertices. Thus restricting the first n subcubes to the first r coordinates does not change which pairs of subcubes intersect. So the intersection graph of these restricted subcubes is G , implying that $G \in I(n, r)$.

Therefore, using the definition of ρ , $\tau(n+1) \geq \tau(G') = r \geq \rho(G) = \rho(n)$.

\square

Combining these two lemmas with the bounds on ρ gives:

Theorem 4.3. *There is some absolute constant, c , such that for all n ,*

$$n - c \log n \leq \tau(n) \leq n - \lfloor \log n \rfloor + 1.$$

4.3 Bounding bp in terms of bc

As the earlier example of K_n shows, $\text{bp}(G)$ can be as large as exponential in $\text{bc}(G)$. Here, we show that $\text{bp}(G)$ grows no faster than exponentially in $\text{bc}(G)$, and calculate the best

upper bound exactly. We do this by reducing the problem to proving the upper bound for a single family of graphs, which we do by induction on graphs in the family. In Theorem 4.6, we use ideas from Tverberg's proof [89] of the Graham-Pollak Theorem to calculate the biclique partition number of these graphs and show that the previous bound is tight.

Theorem 4.4. *If $\text{bc}(G) = m$, then $\text{bp}(G) \leq \frac{1}{2}(3^m - 1)$.*

Proof. Fix an m -cover of G , $\kappa = \{B_1, \dots, B_m\}$, and for each biclique B_i we label the vertex classes as class 0 and class 1. We now represent each vertex of G as an element of $\{0, 1, *\}^m$, based on its membership of elements of κ . More precisely, for $v \in V(G)$, we define $\tilde{v} \in \{0, 1, *\}^m$ by:

$$\tilde{v}_i = \begin{cases} 0 & \text{if } v \text{ is in class 0 of } B_i, \\ 1 & \text{if } v \text{ is in class 1 of } B_i, \\ * & \text{otherwise.} \end{cases}$$

If $\tilde{u} = \tilde{v}$ then u and v have the same neighbours. Therefore, if $u \neq v$, this would mean that $\text{bp}(G) = \text{bp}(G - u)$, so we could replace G with $G - u$ and thus we may assume vertices all have different representations. Using the identification $v \mapsto \tilde{v}$, we have that $V(G) \subseteq \{0, 1, *\}^m$.

Since the biclique partition number of a graph is at least that of each of its induced subgraphs, we may assume that $V(G) = \{0, 1, *\}^m$. Note that there is an edge between two vertices u and v if and only if there is some i for which $\{u_i, v_i\} = \{0, 1\}$. In other words, this is the complement of the intersection graph of all subcubes of an m -dimensional cube. We write G_m for this graph.

Before proceeding by induction on m , we introduce some notation. We write (α, β) for the biclique with vertex classes α and β . If $\alpha \subseteq \{0, 1\}^m$ and $x, y \in \{0, 1, *\}$, we define the following subsets of $\{0, 1, *\}^{m+1}$:

$$x\alpha := \{(x, v_1, v_2, \dots, v_m) : (v_1, \dots, v_m) \in \alpha\} \quad \begin{matrix} x \\ y \end{matrix} \alpha := x\alpha \cup y\alpha$$

Let π_m be a $\text{bp}(G)$ -partition of G_m . Add the biclique of κ , B_1 , to π_{m+1} . For every $(\alpha, \beta) \in \pi_m$, place the following three bicliques in π_{m+1} : $(\begin{smallmatrix} * \\ 0 \end{smallmatrix} \alpha, \begin{smallmatrix} * \\ 0 \end{smallmatrix} \beta)$, $(\begin{smallmatrix} * \\ 1 \end{smallmatrix} \alpha, 1\beta)$ and $(1\alpha, *\beta)$.

Claim 4.5. π_{m+1} is a partition of G_{m+1} and contains $3 \text{bp}(G_m) + 1$ bicliques.

Proof of claim. Let $u = (u_1, \dots, u_{m+1})$ and $v = (v_1, \dots, v_{m+1})$ be two vertices that are not joined by an edge of G_{m+1} . This means there is no i such that $\{u_i, v_i\} = \{0, 1\}$ so uv is not an edge of B_1 . Moreover, if we define $u' = \{u_2, \dots, u_{m+1}\}$ and $v' = \{v_2, \dots, v_{m+1}\}$, i.e. u and v with the first coordinate removed, then $u'v'$ is not an edge of G_m . Hence $u'v'$ is not an edge in any biclique of π_m ; this implies uv is not an edge of any biclique of π_{m+1} .

Conversely, let e be an edge of G_{m+1} . If e is also an edge of B_1 , it is contained in no other bicliques of π_{m+1} . If e is not an edge of B_1 , then removing the first coordinate from each of its endpoints forms an edge e' of G_m . As π_m is a partition of G_m , the edge e' lies in exactly one $(\alpha, \beta) \in \pi_m$. Then, by inspection, e must lie in exactly one of the corresponding bicliques in π_{m+1} . \square

As $\text{bp}(G_1) = 1$, our proof is concluded by applying the above claim inductively. Indeed for any graph G , with $\text{bc}(G) = m$, we have shown that $\text{bp}(G) \leq \text{bp}(G_m) \leq \frac{1}{2}(3^m - 1)$. \square

Theorem 4.6. There is a graph G with $\text{bc}(G) = m$, but $\text{bp}(G) = \frac{1}{2}(3^m - 1)$.

Proof. For a graph G , we write $A(G)$ for the adjacency matrix of the graph, and $\text{rank}(G)$ for the rank of $A(G)$. Suppose P_1, \dots, P_k form a biclique partition of G . Then $A(G) = \sum_{i=1}^k A(P_i)$. Since rank is subadditive and bicliques have rank 2, we have that $\text{rank}(G) \leq 2k$. This implies $\text{bp}(G) \geq \text{rank}(G)/2$.

We conclude the proof by showing that $\text{rank}(G_m) = 3^m - 1$, where G_m is the graph defined in the proof of Theorem 4.4. We define the following order on our symbols: $0 < 1 < *$, and we use lexicographical order for the product spaces $\{0, 1, *\}^m$, for all m . That is, for $u, v \in \{0, 1, *\}^m$, we say $u < v$ if there is some i for which $u_i < v_i$ and $u_j = v_j$ for all $j < i$.

Note that $A_m := A(G_m)$ has 3^m rows (and columns) and that the final row (and column) in the above order corresponds to the vertex $(*, \dots, *)$ and hence consists solely of 0's.

We use the convention that $A_0 = \begin{pmatrix} 0 \end{pmatrix}$, the 1×1 zero matrix, and we write 0_t and 1_t for the $3^t \times 3^t$ all zeros matrix and for the $3^t \times 3^t$ all ones matrix respectively. Then in block

matrix notation,

$$A_m = \begin{pmatrix} A_{m-1} & 1_t & A_{m-1} \\ 1_t & A_{m-1} & A_{m-1} \\ A_{m-1} & A_{m-1} & A_{m-1} \end{pmatrix} \rightarrow \begin{pmatrix} 0_t & 1_t - A_{m-1} & 0_t \\ 1_t - A_{m-1} & 0_t & 0_t \\ A_{m-1} & A_{m-1} & A_{m-1} \end{pmatrix}.$$

The second matrix is obtained from the first by simple row operations. We define A'_m to be the matrix formed by replacing the final row of this second matrix by a row of only 1's. We now use induction on m to show that A'_m is of full rank, that is, the linear span of the rows of A'_m , written $\text{span}(A'_m)$, is \mathbb{R}^{3^m} . Note that

$$A'_m = \begin{pmatrix} 0 & 1 - A_{m-1} & 0 \\ 1 - A_{m-1} & 0 & 0 \\ A'_{m-1} & A'_{m-1} & A'_{m-1} \end{pmatrix}.$$

Since $\text{span}(1 - A_{m-1}) = \text{span}(A'_{m-1}) = \mathbb{R}^{3^{m-1}}$, it is easy to see that $\text{rank}(A'_m) = 3^m$, proving our induction hypothesis. This implies that for all m , $\text{rank}(A_m) = 3^m - 1$ and that $\text{bp}(G_m) \geq \frac{1}{2}(3^m - 1)$. Combined with Theorem 4.4, this concludes the proof. \square

4.4 Relationship between lbc and lbp

As before, we write (α, β) for a biclique with vertex classes α and β .

In contrast to the previous section's results, we show that lbp cannot be bounded by lbc. Indeed, even if $\text{lbc}(G) = 2$, it is possible for $\text{lbp}(G)$ to be arbitrarily large. More concretely:

Theorem 4.7. *For all $m \geq 2$, there is a graph G with a 2-local m -cover for which $\text{lbp}(G) \geq \frac{1}{2} \log(\frac{m-1}{3})$.*

We explicitly construct graphs satisfying this inequality. For $m \geq 2$, let L_m have vertex set

$$V(L_m) = \{v \in \{0, 1, *\}^m \mid v \text{ has exactly } m - 2 \text{ } * \text{'s}\}.$$

We let $uv \in E(L_m)$ if and only if there is some i for which $\{u_i, v_i\} = \{0, 1\}$, where $u = (u_1, \dots, u_m)$ and $v = (v_1, \dots, v_m)$. Although we do not use this fact, L_m is the

complement of the intersection graph of all codimension 2 subcubes of a cube of dimension m . Our definition of L_m may appear strange but it arises from simple reverse engineering of a graph with a 2-local m -cover. See the proof of Theorem 4.10 for more details.

The graph L_m may also be viewed as the union of m bicliques, B_1, \dots, B_m , where B_i is the subgraph induced by the vertex set $\{v \in V(L_m) \mid v_i \neq *\}$. We term these B_i *covering bicliques*, as they form a 2-local cover of L_m . Thus $\text{lbc}(L_m) = 2$.

The *crown graph* on $2t$ vertices, H_t , is $K_{t,t}$ with a perfect matching removed. We may label the vertices of H_t as u^1, \dots, u^t and v^1, \dots, v^t ; hence $u^i v^j$ is an edge if and only if $i \neq j$, and there are no other edges.

We shall show that an r -local partition of L_m can be altered to form an r -local partition of H_t , for t linear in m . We then show that the local biclique cover number of crown graphs tends to infinity as the size of the graph does, indeed, $\text{lbc}(H_t) \geq \frac{1}{2} \lceil \log t \rceil$. Since $\text{lbp} \geq \text{lbc}$, this suffices to finish the proof.

Proof of Theorem 4.7. Let L_m and its subgraphs B_1, \dots, B_m be as above. We say that an edge uv is *shared* if it is an edge in two of the covering bicliques; equivalently there are two i such that $\{u_i, v_i\} = \{0, 1\}$.

Let π be a partition of L_m —we call bicliques in π *partitioning bicliques*. Let e be a shared edge contained in B_i and let $A = (\alpha, \beta)$ be the partitioning biclique containing it. Consider the graph $A \cap B_i$, the graph with vertex set $V(A) \cap V(B_i)$ and edge set $E(A) \cap E(B_i)$. It consists of up to two vertex-disjoint bicliques—let $A_i^e = (\alpha_i^e, \beta_i^e)$ denote the one containing the edge e . We term this the *restriction* of A onto B_i (with respect to the edge e). Note that all vertices in α_i^e agree in the i^{th} coordinate, as do all vertices in β_i^e .

Lemma 4.8. *Let e be a shared edge of B_i and B_j , and $A = (\alpha, \beta)$ a biclique containing it. Then one of A_i^e or A_j^e is a star, or both A_i^e and A_j^e are $K_{2,2}$.*

This lemma, though slightly technical, is key to our approach. A more intuitive formulation is that all bicliques containing an edge shared by two covering bicliques are either ‘close’ to being contained in one of these covering bicliques or ‘very small’.

Proof. By permuting the coordinates and swapping 0 with 1 in the first two coordinates, we may assume that, the endpoints of e are $(0, 0, *, \dots, *)$ and $(1, 1, *, \dots, *)$, with the former

being in α and the latter in β .

Notice that if any of α_1^e , α_2^e , β_1^e or β_2^e are singleton sets, we are done, so we assume the contrary. Suppose at least one of $(0, 1, *, \dots, *)$ and $(1, 0, *, \dots, *)$ is in A —without loss of generality $(0, 1, *, \dots, *)$ is in α . Then every vertex in β must have a 1 in the first coordinate. But any vertex in β_2^e has 1 in the second coordinate. Thus there is only one vertex in β_2^e , contradicting our earlier assumption.

Therefore neither $(0, 1, *, \dots, *)$ nor $(1, 0, *, \dots, *)$ is in A . A second vertex in α_1^e must have a 0 in the first coordinate, a $*$ in the second coordinate and without loss of generality, a 0 in the third coordinate. Equally, a second vertex in β_2^e must be $(*, 1, 1, *, \dots, *)$ in order to be adjacent to the vertices we have assumed are in α .

A further vertex in α_2^e must start with $*$, followed by 0 and have a further non-asterisk digit, in the i^{th} place, say, $(i > 2)$. Our previous assumptions determine that there is only one further vertex in A —the vector with 1's in the first and i^{th} coordinates. It can be seen that A_1 and A_2 are both $K_{2,2}$.

□

We now define for each i , a colouring c_i of the shared edges of B_i . If e is a shared edge of B_i and B_j and A is the partitioning biclique containing it, $c_i(e)$ is red if A_i^e is a star or a $K_{2,2}$ and blue otherwise. Note that the preceding lemma shows that if $c_i(e)$ is blue then A_j^e is a star.

A shared edge is blue in at most one colouring. Hence, each of the $2\binom{m}{2} = m(m-1)$ shared edges is red in at least one of the m colourings, implying that at least one of the m covering bicliques, B_1 say, must contain at least $m-1$ shared edges coloured red.

All shared edges are vertex-disjoint—as seen by the form of the edges as vectors—so we may label the edges as $u^1v^1, \dots, u^{m-1}v^{m-1}$, where the u^i and the v^j are from different classes of B_1 . We seek a large $B \subseteq B_1$ such that all shared edges in B are in partitioning bicliques that restrict to stars on B_1 . Pick $I \subseteq [m-1]$ as follows. Let $1 \in I$. If the shared edge $e = u^1v^1 \in A \in \pi$ and $A_1^e \cong K_{2,2}$ then discard from I the (at most two) indices other than 1 for which one of u^i and v^i is a vertex of this $K_{2,2}$. Then proceed to the next index not already discarded, and continue in the same way. Since at each stage, we throw away at most two indices, the set of surviving indices, I , satisfies $|I| \geq (m-1)/3$.

Let B be the induced subgraph of B_1 with the u^i and v^i as vertices, for $i \in I$. Let H be the subgraph of B formed by removing all shared edges, $u^i v^i$ for $i \in I$. Clearly, H is isomorphic to $H_{|I|}$, the crown graph on $2|I|$ vertices. Let A be a biclique containing one of these removed edges, e , and note that $A \cap B$ is the union of at most two disjoint bicliques. By the construction of B , the component of $A \cap B$ containing e , i.e. $A_1^e \cap B$, is a star. We label the other (possibly empty) biclique component A' . From π , we induce a partition π' of H as follows. If $A \in \pi$ contains an edge of H , then place $A \cap H$ inside π' . Now, $A \cap H = A \cap B$ and is hence a union of bicliques, unless A contains a removed edge. Using $G - e$ to denote the graph G with the edge e deleted, in this case, $A \cap H = A' \cup (A_1^e \cap B - e)$, the latter component remaining a star. Note that if π is a k -local partition of L_m , then π' is a k -local partition of H . This argument is valid for all partitions π of L_m , so $\text{lbp}(H) \leq \text{lbp}(L_m)$.

We now complete the proof in the following lemma:

Lemma 4.9. $\text{lbc}(H_t) \geq \frac{1}{2} \lceil \log t \rceil$.

Proof. A biclique cover of H_t induces a biclique cover of K_t when each pair of corresponding vertices (i.e. u^i and v^i for each i) are identified as a single vertex. As mentioned earlier, $\text{lbc}(K_t) = \lceil \log t \rceil$ so, there must be some vertex of the K_t in at least $\lceil \log t \rceil$ bicliques belonging to the induced cover. This implies one of the vertices of H_t that are identified to it is in at least $\frac{1}{2} \lceil \log t \rceil$ bicliques, in the original covering. □

Hence we have shown that, $\text{lbp}(L_m) \geq \text{lbp}(H_{(m-1)/3}) \geq \frac{1}{2} \log \frac{m-1}{3}$, which concludes the proof. □

4.5 Bounding $\text{lbp}(G)$, for $\text{lbc}(G) = 2$

The previous section showed that even for $\text{lbc}(G) = 2$, we cannot place an absolute bound on $\text{lbp}(G)$. In this section, we prove a bound on lbp that makes use of the number of bicliques in a 2-local cover of G .

Theorem 4.10. *If G has a 2-local m -cover then $\text{lbp}(G) \leq \lceil \log(m-1) \rceil + 3$.*

By Theorem 4.7, this upper bound is best possible up to a constant factor.

We first prove an upper bound on the local bipartite partition number of crown graphs, which is a key element of the proof of the theorem. We then proceed by using an argument similar to the start of the proof of Theorem 4.4 to reduce proving the upper bound in the theorem for all graphs with a 2-local m -cover to proving it for one particular graph. This ‘worst’ graph will turn out to be similar to the graph used in the proof of Theorem 4.7.

Lemma 4.11. $\text{lb}_p(H_t) \leq \lceil \log t \rceil$.

Proof. Let π be a $\lceil \log t \rceil$ -local partition of K_t , whose vertices we label x_1, x_2, \dots, x_t . Such a partition is fairly easy to construct; see for instance [29]. Given a set $\alpha \subseteq V(K_t)$, we define $u^\alpha := \{u^i : x_i \in \alpha\}$ and $v^\alpha := \{v^i : x_i \in \alpha\}$. Let π' be the collection of bicliques of the form (u^α, v^β) or (u^β, v^α) , for $(\alpha, \beta) \in \pi$. It is easy to see π' is a $\lceil \log t \rceil$ -partition of H_t . \square

Proof of Theorem 4.10. Fix a 2-local m -cover of G . We may identify vertices $v \in V(G)$ with vectors $\tilde{v} \in \{0, 1, *\}^m$ with exactly $m - 2$ $*$ ’s, similar to the proof of Theorem 4.4. As shown there, we may assume no two vertices have the same representation, and that all such vectors are the representation of some vertex, i.e. we assume that

$$V(G) = \{v \in \{0, 1, *\}^m \mid v \text{ has at least } m - 2 \text{ } * \text{'s}\}.$$

Note that there is an edge between two vertices u and v if and only if there is some i for which $\{u_i, v_i\} = \{0, 1\}$.

For $i = 1, \dots, m$, let B_i be the induced subgraph on the vertices whose i^{th} coordinate is not $*$. We define shared edges in an identical manner to the proof of Theorem 4.7 and we shall define C_i , a subgraph of B_i , such that every edge of G is an edge in exactly one C_i .

Let C_i contain all non-shared edges of B_i . Additionally, for all i and j , let C_i contain exactly one of the edges shared between B_i and B_j , and let C_j contain the other. Since C_i is a complete bipartite graph minus a matching of size $m - 1$, it contains an induced subgraph isomorphic to H_{m-1} . Thus by Lemma 4.11, we may partition each C_i into bicliques in such a way that every vertex in the vertex set of this H_{m-1} lies in at most $\lceil \log(m - 1) \rceil + 1$ bicliques and every other vertex lies in at most 2 bicliques. Clearly, every vertex of G lies in at most two of the C_i and in the partition of one of them, it lies in at most 2 bicliques. This observation concludes the proof. \square

4.6 Discussion

Since $\text{bc}(G)$ is a measure of independent interest to local measures, one may try to use the separate existence of a m -cover in a graph with $\text{lbc}(G) = 2$, rather than the existence of a 2-local m -cover, to bound $\text{lbp}(G)$. More formally, we ask:

Question 4.1. *What is the smallest $k(m)$ such that if the graph G has $\text{bc}(G) = m$ and $\text{lbc}(G) = 2$, we have $\text{lbp}(G) \leq k$?*

Theorem 4.7 tells us that $k(m) \geq \frac{1+o(1)}{2} \log m$. However, an upper bound does not follow directly from Theorem 4.10 as the cover that shows $\text{lbc}(G) = 2$ may be genuinely different from the cover that shows $\text{bc}(G) = m$. The following result shows this:

Theorem 4.12. *For all $m \geq 4$, there is a bipartite graph, G , with $\text{lbp}(G) = \text{lbc}(G) = 2$ and $\text{bp}(G) = \text{bc}(G) = m$, but there are no $(m-1)$ -local m -covers.*

Proof. First, we make a general observation about biclique covers. Two edges, $\{v_1, v_2\}$ and $\{u_1, u_2\} \in E(G)$ are called *strongly independent* if they are independent and the minimum degree of $G[u_1, u_2, v_1, v_2]$ is 1—in other words, they are vertex disjoint and do not lie in the same $K_{2,2}$ subgraph. No biclique can contain both edges so if $E(G)$ contains a set of k pairwise strongly independent edges, $\text{bc}(G) \geq k$.

We define a bipartite graph G as follows. We let $V(G) = X \cup Y$, where $X = \{x_1, \dots, x_m\}$ and $Y = \{y_{i,j} : 1 \leq i < j \leq m\} \cup \{y_{[m]}\}$. Let there be an edge from x_i to $y_{i,j}$ and to $y_{j,i}$ for all appropriate j . Also, let $y_{[m]}$ have edges to all of the x_i .

Consider the partition π consisting of a $K_{1,m}$ with its centre at $y_{[m]}$ and stars at each of the x_i each with leaves at all adjacent vertices other than $y_{[m]}$. Thus $\text{lbp}(G) = \text{lbc}(G) = 2$. Note that $I := \{\{x_1, y_{1,2}\}, \{x_2, y_{2,3}\}, \dots, \{x_{m-1}, y_{m-1,m}\} \cup \{x_m, y_{1,m}\}$ is a set of pairwise strongly independent edges. Therefore $\text{bc}(G) \geq m$, and so $\text{bp}(G) = \text{bc}(G) = m$ as we can cover (indeed partition) with m stars, one centred at each x_i .

Suppose κ is a biclique cover of G using just m bicliques. Each of these contains exactly one of the strongly independent edges listed above.

We can see that $I \cup \{x_1, y_{1,3}\} \setminus \{x_1, y_{1,2}\}$ is also a strongly independent set of m edges. Thus the biclique in κ that contains the edge $\{x_1, y_{1,2}\}$ must also contain the edge $\{x_1, y_{1,3}\}$.

Since x_1 is the only common neighbour of $y_{1,2}$ and $y_{1,3}$, the vertex x_1 must be at the centre of a star in κ . By symmetry of the x_i , κ must contain m stars, one centred at each x_i . Since these account for all the bicliques in κ , the vertex $y_{[m]}$ must lie in each star. Hence κ is not $(m-1)$ -local.

□

So we can see that insisting that a cover of a graph G attains $\text{bc}(G)$ may ensure the cover is not close to being optimal for lbc . This leads to the following question—if we insist that the cover of a graph attains $\text{lbc}(G)$, could this ensure that it is far from being optimal for bc ? More formally:

Question 4.2. *Suppose $\text{bc}(G) = m$, $\text{lbc}(G) = k$, what is the smallest $r = r(m, k)$ such that we can guarantee G has a k -local r -cover?*

An answer to this for $k = 2$, combined with Theorem 4.10, would lead to an upper bound for Question 4.1.

The following example of the hypercube is instructive, in giving bounds on the problem. In particular, it shows $r(2^{d-1}, d/2) \geq d2^{d-3}$, for d even.

Example. *Consider Q_d , for d even. It is easy to show that $\text{bp}(Q_d) = \text{bc}(Q_d) = 2^{d-1}$. Indeed, each biclique contains at most d edges, and $e(Q_d) = d2^{d-1}$, so $\text{bc}(Q_d) \geq 2^{d-1}$. On the other hand, we may partition using $|Q_d|/2 = 2^{d-1}$ stars since Q_d is bipartite. Dong and Liu [29] showed that $\text{lbp}(Q_d) = \text{lbc}(Q_d) = d/2$. In fact, their proof showed that a cover achieves this only if it consists solely of $K_{2,2}$'s so every $d/2$ -local cover contains exactly $d2^{d-3}$ bicliques.*

4.6.1 Complexity of Boolean Matrices

Here we discuss a notion closely related to bipartite covers and discuss its relationship to a model of the computational complexity of linear operators. We then discuss a use, by Jukna and Sergeev [54], of the proof of Theorem 4.7 in this context.

An $n \times n$ matrix, R , is called a *rectangle* if there are sets $I, J \subseteq [n]$ such that $R_{ij} = 1$ if $i \in I$ and $j \in J$, with $R_{ij} = 0$ otherwise. A *rectangle cover* of a $(0,1)$ square matrix, A , is a set of rectangles such that $A_{ij} = 0$ if and only if $R_{ij} = 0$ for each rectangle R . A *rectangle*

partition of a $(0, 1)$ -matrix A is a set of rectangles R^1, \dots, R^k such that $A = \sum_i R^i$. Clearly, all rectangle partitions are also rectangle covers. The *weight* of a rectangle cover is the total number of non-zero rows and columns in the rectangles. Since the adjacency matrix of a biclique is the sum of two rectangles, a biclique cover of a graph is naturally associated with a rectangle cover of its adjacency matrix. The weight of a biclique cover of G is twice the weight of the associated rectangle cover of its adjacency matrix.

Let $(S, +)$ be an Abelian additive semigroup, that is, S is a set and $+: S \times S \rightarrow S$ is commutative and associative. We also let A be an $n \times n$ $(0, 1)$ -matrix. The goal of this branch of research is to compute the equation $y = Ax$, over the semigroup S , using what is known as an *addition circuit*. Since 0 and 1 may not be in S , and we do not necessarily have a concept of multiplication in S , we interpret the equation as representing the system of sums $y_i = \sum_{j: a_{ij}=1} x_j$. This consists of a directed acyclic graph with n sources, nodes with zero indegree, and n sinks, nodes with zero outdegree. We associate the source nodes with the n inputs, x_1, \dots, x_n , and the output nodes with the outputs $y_1 \dots y_n$ to each of the outputs we associate. Every node that is not a source computes the sum (carried out in the semigroup $(S, +)$) of its the vertices in its in-neighbourhood. We say the circuit computes A over $(S, +)$ if the equation $y = Ax$ holds for all inputs $x = (x_1, \dots, x_n)$. See also the survey of Jukna and Sergeev [54] for a more detailed introduction to the area.

The two most commonly studied semigroups, are the SUM semigroup, which is \mathbb{N} together with the usual $+$ operator, as well as the OR semigroup, $(\{0, 1\}, \vee)$, where $a \vee b = 0$ if and only if $a = b = 0$.

A circuit is said to have depth d if every path from source to sink has d edges. One of the notions of complexity of a matrix A , written $\text{SUM}_d(A)$, is the minimum number of edges in a depth d circuit that computes A over SUM. Similarly $\text{OR}_d(A)$ is the least number of edges in a depth d circuit that computes A over OR.

In the case where $d = 2$, there is a close connection between the circuit complexity and rectangle covers of matrices. Indeed, $\text{OR}_2(A)$ is the lowest weight of a rectangle cover of A and $\text{SUM}_2(A)$ is the lowest weight of a rectangle partition of A .

Since the publication of Theorem 4.7 in [81], its proof has been translated into this language, by Jukna and Sergeev [54], to prove that there are $n \times n$ matrices, M_n , such that

$\text{OR}_2(M_n) \leq 4n$ but $\text{SUM}_2(M_n) \geq n \log n$, thus demonstrating a significant gap is possible between OR_2 and SUM_2 .

Chapter 5

Uniformly Random Simplicial Complexes

5.1 Introduction

A *simplicial complex on n vertices*, Δ , is a collection of non-empty subsets of $[n] := \{1, \dots, n\}$ such that if $x \in \Delta$ and $\emptyset \neq y \subseteq x$ then $y \in \Delta$. Note that simplicial complexes are closely related to down sets; indeed, Δ is a simplicial complex if and only if $\Delta \cup \{\emptyset\}$ is a down set. In the literature, simplicial complexes are sometimes termed *abstract simplicial complexes*, to distinguish them from a closely related object, known as a geometric simplicial complex, which we do not discuss here.

We refer to the elements of a simplicial complex as *faces* and call the faces of size one *vertices*. A *facet* is a maximal face, that is, a face that is not contained in another face of the simplicial complex. The *dimension* of a face x is $|x| - 1$. In other words it is one less than the size of x as a subset of $[n]$. A *pure* simplicial complex is a simplicial complex whose facets all have the same dimension .

The t^{th} *face number* of Δ , written $f_t = f_t(\Delta)$, is the number of faces of size t . Note that in the literature f_t is sometimes instead defined as the number of faces of dimension t . We write $[n]^{(t)}$ for the collection of size t subsets of $[n]$. The t -*skeleton* of Δ , written $\Delta^{(t)}$, is the complex formed from the collection of faces of Δ of size at most t . We say Δ has a complete t -skeleton if $[n]^{(t)} \subseteq \Delta^{(t)}$. Equivalently, Δ has a complete t -skeleton if $f_t(\Delta) = \binom{n}{t}$.

Simplicial complexes are a long-studied mathematical object. In recent years, random simplicial complexes have received much attention, in particular the model introduced by Linial and Meshulam in [66]. For a fixed constant, t , they generate simplicial complexes on n vertices by starting with a complete $(t - 1)$ -skeleton and adding t -sets uniformly independently at random, with probability p . We write $\Delta \sim \mathcal{LM}(n, t, p)$ if Δ is generated in this manner. Topological properties of this model has been studied further by many authors, for instance see [73], [4], [6] and [40]. For ease of reference, this model and all other random models mentioned in this paper are listed in a brief summary at the end of this introduction—see Section 5.1.2.

Another model in the literature is the frame-wise uniform model of Brooke-Taylor and Testa [19], in which they build simplicial complexes on n vertices by adding faces inductively by dimension. More precisely, they first add the subsets of $[n]$ of size 2 with probability $\frac{1}{2}$, and then in the i^{th} stage they consider sets of dimension i (i.e. size $i + 1$) whose non-empty

proper subsets have all already been added. They add such sets with probability $\frac{1}{2}$. We stop this process in the j^{th} stage if all sets of dimension j have some non-empty proper subset which has not been added to the complex.

However, not much work has been done on uniformly random simplicial complexes, despite this being an incredibly natural model. We write $\mathcal{M}(n)$ for the collection of simplicial complexes on n variables, and write $\mathcal{U}(n)$ for the uniform distribution on $\mathcal{M}(n)$. Perhaps the reason for the lack of attention this model has received is the awkwardness of generating a uniformly random simplicial complex. Using a powerful combinatorial characterization due to Korshunov [59] of almost all simplicial complexes on n vertices, see Section 5.1.1, we exhibit a method for working with this uniform model. In this chapter, we demonstrate the strength of this approach by proving various topological properties of this model.

We also study another model which has received limited attention, a random pure model with facets of size t , which we write $\mathcal{RP}(n, t, p)$. We generate $\Delta \sim \mathcal{RP}(n, t, p)$ by choosing subsets of $[n]$ of size t to be faces of Δ uniformly, independently with probability p . We also add to Δ all non-empty subsets of these sets.

We again pay attention to mainly topological properties of this model, which we study in two main ranges: when $t = cn$, for some constant $0 < c < 1$, and when t is a constant. The first of these ranges is of interest for its connection to the uniform model. Indeed, we will see that $\mathcal{RP}(n, \frac{n}{2}, \frac{1}{2})$ is a useful heuristic for $\mathcal{U}(n)$, when n is even. Our interest in the latter range stems mainly from its connection to the Linial–Meshulam model; we see later that there is a range of p for which the models agree, for high probability events. When $t = 2$, something stronger is trivially true: the two models are equivalent to each other and also to the famous Erdős–Renyi random graph, $G(n, p)$.

Much of the work on the Linial–Meshulam model has focused on its homology groups, a set of isomorphism invariants of simplicial complexes. Let \mathbb{F} be a field. The k^{th} *chain group* of Δ over \mathbb{F} , written $K_k(\Delta, \mathbb{F})$, is the set of formal \mathbb{F} -linear combinations of k -dimensional faces of Δ . Thus $K_k(\Delta, \mathbb{F})$ is a vector space with dimension $f_{k+1}(\Delta)$ and has as a basis the collection of faces of dimension k , i.e. of size $k + 1$.

The *boundary map* is the linear map $d_k : K_k(\Delta, \mathbb{F}) \rightarrow K_{k-1}(\Delta, \mathbb{F})$ defined by:

$$d_k(\{x_0, \dots, x_k\}) = \sum_{i=0}^k (-1)^i \{x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_k\},$$

where $x_0 < x_1 < \dots < x_k$ are elements of $[n]$. An important basic property of these maps is that the image of d_{k+1} is contained in the kernel of d_k . In other words, the composite map $d_k \circ d_{k+1} : K_{k+1} \rightarrow K_{k-1}$ is the zero map.

The k^{th} *homology group* of Δ over \mathbb{F} , written $H_k(\Delta, \mathbb{F})$, is defined by the quotient,

$$H_k(\Delta, \mathbb{F}) = \ker(d_k) / \text{im}(d_{k+1}).$$

The majority of the work that has been done on the Linial and Meshulam model has been studying its highest dimension homology group. The latest result on these lines is the following:

Theorem 5.1 (Aronshtam, Linial [5]). *Let \mathbb{F} be a field and let $\Delta \sim \mathcal{LM}(n, t, p)$. With high probability, $H_{t-1}(\Delta, \mathbb{F})$ is non-trivial if $p \geq c_t/n$, where $c_t = t - \frac{t^2-t+1}{\exp(t)} + O\left(\frac{t^2}{\exp(2t)}\right)$.*

In Section 5.4, we investigate the homology group of uniformly random simplicial complexes, in particular showing the following:

Theorem 5.2. *Let \mathbb{F} be a field and let $\Delta \sim \mathcal{U}(n)$, for even n . With high probability, $H_{n/2-1}(\Delta, \mathbb{F})$ has dimension at least $(1 + o(1)) \frac{2^{(n-1)/2}}{\sqrt{\pi n}}$.*

We also prove a version of this for odd n , as well as proving a version of Theorem 5.1 for the random pure complex. More precisely, when t is a constant and when $t = cn$ for some constant c , we find a value $p_0(t)$ such that if $t > p_0(t)$ and $\Delta \sim \mathcal{RP}(n, t, p)$ then with high probability $H_{t-1}(\Delta, \mathbb{F})$ is non-trivial.

Another important isomorphism invariant of simplicial complexes, Δ , is the *Euler characteristic*, written $\chi(\Delta)$, which is defined by:

$$\chi(\Delta) = \sum_{i=1}^n (-1)^{i-1} f_i.$$

In Section 5.5, we prove high probability bounds on the Euler characteristic of the uniform model for even n .

Theorem 5.3. *Let $\Delta \sim \mathcal{U}(n)$, for even n . There is a constant, c , such that with high probability, $|\chi(\Delta)| \leq cn2^{n/2}$.*

In the case of odd n , we also bound the Euler characteristic of a uniform complex to within an interval of width $O(n2^{n/2})$, but we find this interval does not include zero.

Section 5.3 contains work on the skeletons of our models. We show that with high probability, $\Delta \sim \mathcal{U}(n)$ has a complete $(\frac{n}{2} - 2)$ -skeleton, when n is even. We also find ranges of p that guarantee, with high probability, complete t' -skeletons for $\Delta \sim \mathcal{RP}(n, t, p)$, when t' is close to t . We believe these results to be of independent interest, but we also require them in the remainder of the chapter.

Both for the homology work and the work on the skeletons, we find that our results agree with the claim that for even n , the models $\mathcal{U}(n)$ and $\mathcal{RP}(n, \frac{n}{2}, \frac{1}{2})$ behave similarly. The results can also be proved with similar methods (although when working with the skeletons of the uniform model we use a short-cut). We demonstrate an instance when the two models appear very different to work with, in Section 5.6, when we investigate the property of shellability for both models; we recall the definition of this concept in that section.

The final part of this chapter, Section 5.7, deals briefly with our original motivation for studying the uniformly random complex, the notion of evasiveness. A *Boolean function* on n variables is a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$. Informally, the *decision tree complexity* of f , written $D(f)$, is the minimum number of bits of input an adaptive algorithm, which knows the function, must know in order to determine the output. More precisely, the algorithm initially has no information about the input $x = (x_1, \dots, x_n)$ and at each stage asks a question of the form ‘what is x_i ?’, where its choice of question may depend on previously received answers and where the algorithm knows the value for the function for every possible input. The algorithm terminates when it knows the value of $f(x)$. We say $D(f) \leq t$, if there is some algorithm that terminates after at most t questions for all inputs x . Trivially, for all functions f with n variables, $D(f) \leq n$. We say f is *evasive* if $D(f) = n$. For example, it is easy to see that the projection functions, $p_k(x_1, \dots, x_n) = x_k$, are not evasive (for $n > 1$), and that the layer functions, defined by $l_k(x_1, \dots, x_n) = 1$ if and only if $\sum_{i=1}^n x_i = k$, are evasive.

In 1976, Rivest and Vuillemin [84] showed that almost all Boolean functions are evasive,

in the sense that the proportion of functions on n variables that are evasive tends to 1, as $n \rightarrow \infty$. Given this result, it is natural to seek a large class of Boolean functions, all of which are evasive.

A *symmetry* of a Boolean function f is a permutation $\phi : [n] \rightarrow [n]$ such that $f(x_1, \dots, x_n) = f(x_{\phi(1)}, \dots, x_{\phi(n)})$, for all tuples (x_1, \dots, x_n) . We say f is *transitive* if it has a transitive group of symmetries, i.e., for all i and j , there is some symmetry ϕ such that $\phi(i) = j$. A Boolean function, f , is *monotone* if $f(x_1, \dots, x_n) \leq f(y_1, \dots, y_n)$, whenever $x_i \leq y_i$ for all i .

Much of the work in the area of evasiveness (see for instance [70], [64]) is around the following conjecture of Kahn, Saks and Sturtevant, posed in [55].

Conjecture 5.1 (Evasiveness Conjecture [55]). *Every non-constant monotone, transitive Boolean function is evasive.*

If the condition of transitivity is relaxed, the function need not be evasive, as demonstrated by the projection functions as shown above. However, the following easy corollary of our work with the uniformly random simplicial complex shows that this is the exception rather than the rule. More precisely, we have the following corollary to Theorem 5.2 and its analogue for odd n .

Corollary 5.4. *Almost all monotone Boolean functions are evasive.*

5.1.1 Korshunov's characterization of simplicial complexes

To introduce Korshunov's result, we require a bit of terminology. We decompose a family of k -sets A into *bundles* by declaring two sets, x_1 and x_2 to be in the same bundle if there is a sequence of k -sets, y_0, \dots, y_t such that $y_0 = x_1$, $y_t = x_2$ and for each i , $|y_i \cap y_{i+1}| = k - 1$.

Definition 5.1. *Let n and t be natural numbers with $t \leq n$. We say a pair of set families, (A, B) is an admissible pair if all of the following hold:*

1. $A \subseteq [n]^{(t-1)}$ and $B \subseteq [n]^{(t+1)}$.
2. $|A| \leq 2^{n/2}$.
3. $|B| \leq 2^{n/2}$.

4. For all $a \in A$ and $b \in B$, we have that $a \not\subset b$.

5. A and B both consist of one and two element bundles, and each have at most n^4 two-element bundles.

For admissible pairs (A, B) , we write $\mathcal{M}(n, t, A, B)$ for the collection of simplicial complexes on n vertices such that A is the set of facets of size $t - 1$, B is the set of facets of size $t + 1$ and all the other facets have size t . Let also $\mathcal{M}(n, t)$ be the union over admissible (A, B) of $\mathcal{M}(n, t, A, B)$.

We write $\mathcal{M}(n)$ for the collection of simplicial complexes on n vertices. The following result of Korshunov [59] (also stated in the more accessible [60]) .

Theorem 5.5 (Korshunov [59]). *If n is even, then*

$$|\mathcal{M}(n)| = (1 + o(1))|\mathcal{M}(n, n/2)|.$$

If instead n is odd, then

$$|\mathcal{M}(n)| = (1 + o(1))(|\mathcal{M}(n, \lfloor n/2 \rfloor)| + |\mathcal{M}(n, \lceil n/2 \rceil)|).$$

We note briefly that there is a slight asymmetry in the definition of $\mathcal{M}(n, n/2)$. Each complex in $\mathcal{M}(n, n/2)$ does not contain any face of size greater than $n/2 + 1$ but it need not contain every face of size less than $n/2 - 1$. However, the proportion of complexes in $\mathcal{M}(n, n/2)$ that does not contain every face of size less than $n/2 - 1$ is asymptotically zero.

We say an admissible pair (A, B) is *strongly admissible* if $\left| |A| - \binom{n}{n/2+1} 2^{-n/2-1} \right| \leq n 2^{n/4}$ and $\left| |B| - \binom{n}{n/2+1} 2^{-n/2-1} \right| \leq n 2^{n/4}$. The following result, proved in [59] is stated in [61] by Korshunov and Shmulevich.

Theorem 5.6 (Korshunov [59]). *If n is even, then almost all simplicial complexes are strongly admissible.*

We write $\mathcal{U}(n, t, A, B)$ for the uniform distribution on elements of $\mathcal{M}(n, t, A, B)$ and similarly define $\mathcal{U}(n)$ and $\mathcal{U}(n, t)$.

Korshunov's theorems are key to our work on the uniform random simplicial complex. In particular, they allow us to prove high probability results for $\mathcal{U}(n)$ by proving these results hold with high probability for $\mathcal{U}(n, t, A, B)$ for $t \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$ and all admissible (A, B) .

An element of $\mathcal{M}(n, t, A, B)$ is determined by its facets of size t . Let F be the collection of all t -sets which contain no set $a \in A$ and are not contained in any $b \in B$. We call such set *free sets*. When $t \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$, the bounds on A and B imply that $\binom{n}{n/2} - (n/2 + 1)2^{n/2+1} \leq |F| \leq \binom{n}{n/2}$; in particular, there are $(1 + o(1))\binom{n}{t}$ free sets. Picking a uniformly random element, Δ , of $\mathcal{M}(n, t, A, B)$ is equivalent to picking free sets to be faces of the complex uniformly, independently at random with probability $\frac{1}{2}$. This observation makes the distributions $\mathcal{U}(n, t, A, B)$ much easier to work with than working with $\mathcal{U}(n)$ directly.

Given the form of Theorem 5.5, it is perhaps unsurprising that the behaviour of $\mathcal{U}(n)$ depends on the parity of n , as alluded to above. We work mainly with even n as the results are cleaner to state.

We note briefly that Theorems 5.5 and 5.6 were proved on the way to the following asymptotic solution to Dedekind's problem, the problem of enumerating monotone Boolean functions (this is equivalent to enumerating simplicial complexes).

Theorem 5.7 (Korshunov [59]). *If n is even,*

$$|\mathcal{M}(n)| = (1 + o(1))2^{\binom{n}{n/2}} \exp \left[\binom{n}{n/2 - 1} \left(2^{-\frac{n}{2}} + n^2 2^{-n-5} - n 2^{-n-4} \right) \right].$$

If n is odd,

$$|\mathcal{M}(n)| = (1 + o(1))2^{\binom{n}{\lfloor n/2 \rfloor} + 1} \exp \left[\binom{n}{\lfloor n/2 \rfloor - 1} \left(2^{-\frac{n+3}{2}} + n^2 2^{-n-6} - n 2^{-n-3} \right) + \binom{n}{\lfloor n/2 \rfloor} \left(2^{-\lceil n/2 \rceil} + n^2 2^{-n-4} \right) \right].$$

5.1.2 Summary of models

Since we mention a large number of random models of simplicial complexes in this chapter, we include a brief summary list here. The two that we study in this chapter are the random pure model and the uniformly random model.

- **Frame-wise uniform model.** Simplicial complexes on n vertices are built inductively by adding faces inductively by dimension, adding a face with probability $\frac{1}{2}$ if all its subfaces have already been added.

- Linial–Meshulam model, $\mathcal{LM}(n, t, p)$. Simplicial complexes are generated by starting with a complete $(t - 1)$ -skeleton on n vertices and adding sets of size t uniformly, independently at random, with probability p .
- Random pure model, $\mathcal{RP}(n, t, p)$. Sets in $[n]^{(t)}$ are chosen to be facets uniformly independently at random, with probability p .
- Uniformly random model, $\mathcal{U}(n)$. A simplicial complex is chosen uniformly at random from the collection of all simplicial complexes on n vertices. In this chapter, we use the related models $\mathcal{U}(n, t, A, B)$ and $\mathcal{U}(n, t)$ as an intermediate step in studying $\mathcal{U}(n)$. See Section 5.1.1 for more details on these auxiliary models.

5.2 Preliminaries

5.2.1 The Probabilistic Method

In this paper, we use several results from probability and probabilistic combinatorics. Although the results are quite basic, we state them here and refer the reader to [3] for a fuller introduction.

Theorem 5.8 (Markov’s Inequality). *Let X be a non-negative random variable. Then for all k ,*

$$\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}(X)}{k}.$$

We will use this result several times to bound the probability of a random variable being much larger than its mean. It also has the following useful but easy corollary:

Theorem 5.9 (Chebyshev’s Inequality). *Let X be a random variable, and let $\text{Var}(X) = \sigma^2$. Then for any $\lambda > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

In much of what follows, we are interested in $\mathbb{Z}_{\geq 0}$ -valued random variables, X , that depend on a parameter, n , that represents the number of vertices of a random simplicial complex, and we will be interested in the behaviour of X as $n \rightarrow \infty$. The $k = 1$ case of Markov’s Inequality implies that if $\mathbb{E}(X) \rightarrow 0$, then with high probability, $X = 0$.

On the other hand, if $\mathbb{E}(X) \rightarrow \infty$ and $\text{Var}(X) = o(\exp(X)^2)$, then Chebyshev's Inequality allows us to deduce that with high probability, $X = (1 + o(1))\mathbb{E}(X)$. We use both these inequalities frequently in what follows.

We will also make frequent use of the following tool which allows us to calculate variances, usually for the purpose of applying Chebyshev's Inequality.

Lemma 5.10. *Suppose X_1, \dots, X_m are indicator random variables. Then*

$$\text{Var}(X) \leq \mathbb{E}(X) + \sum_{i \neq j} \text{Cov}(X_i, X_j).$$

When we know that a variable is binomially distributed, we have an alternative to using Chebyshev's Inequality that gives much stronger bounds.

Theorem 5.11 (Hoeffding's inequality [47]). *Let X be a binomial random variable with parameters n and p , and let $0 < \epsilon < \min(p, 1 - p)$. Then*

$$\mathbb{P}(|X - pn| \geq \epsilon n) \leq \exp(-2\epsilon^2 n).$$

5.2.2 Asymptotic Results

We require the following of asymptotic results for estimating factorials and binomial coefficients.

Proposition 5.12 (Stirling's formula). *As n tends to infinity,*

$$\log(n!) = n \log(n) - n + \frac{1}{2} \log(2\pi n) + o(1).$$

Equivalently,

$$n! = (1 + o(1)) \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

From the above, we may prove the following useful approximation for binomial coefficients. We write $\alpha(c) = -c \log(c) - (1 - c) \log(1 - c)$.

Proposition 5.13. *If c is a constant and $0 < c < 1$, then, as n tends to infinity,*

$$\log \binom{n}{cn} = \alpha(c)n + O(\log n).$$

We may deduce this easily from Stirling's formula.

Proof. By applying Stirling's formula to the three factorials in the definition of the binomial coefficient, we have:

$$\begin{aligned}
\log \binom{n}{cn} &= n \log n - n + \frac{1}{2} \log(2\pi n) \\
&\quad - cn \log(cn) + cn - \frac{1}{2} \log(2c\pi n) \\
&\quad - (1-c)n \log((1-c)n) + (1-c)n - \frac{1}{2} \log(2(1-c)\pi n) \\
&= n \log n - cn \log n - cn \log c - (1-c)n \log n - (1-c) \log(1-c) + O(\log n) \\
&= (-c \log c - (1-c) \log(1-c))n + O(\log n).
\end{aligned}$$

□

5.3 Complete Skeletons

In this section, we investigate the skeletons of our models. We start by considering the uniform random complex, for even n , exploiting a symmetry of the distribution $\mathcal{U}(n)$ to deduce from Theorem 5.5 the following statement.

Proposition 5.14. *Suppose $\Delta \sim \mathcal{U}(n)$, where n is even. Then with high probability, Δ has a complete $(\frac{n}{2} - 2)$ -skeleton.*

Proof. Given $\Delta \sim \mathcal{U}(n)$, we define Δ^c by the relation $x \notin \Delta \Rightarrow [n] \setminus x \in \Delta^c$. Note that $\Delta^c \sim \mathcal{U}(n)$ and hence, by Theorem 5.5, it contains no faces of size $\frac{n}{2} + 2$ with high probability. Thus Δ has a complete $(\frac{n}{2} - 2)$ -skeleton with high probability. □

The situation is more complicated when n is odd, but we can use a similar method to prove the following.

Proposition 5.15. *Let n be odd. If $\Delta \sim \mathcal{U}(n, \lceil n/2 \rceil)$, then with high probability, Δ has a complete $(\lceil n/2 \rceil - 2)$ -skeleton. If $\Delta \sim \mathcal{U}(n, \lfloor n/2 \rfloor)$, then with high probability, Δ has a complete $(\lfloor n/2 \rfloor - 2)$ -skeleton.*

Proof. Let $\Delta \sim \mathcal{U}(n, \lceil n/2 \rceil)$. It follows easily from Theorem 5.5 that, with high probability, $\Delta^c \in \mathcal{M}(n, \lceil n/2 \rceil)$ or $\Delta^c \in \mathcal{M}(n, \lfloor n/2 \rfloor)$. But since $\Delta^c \in \mathcal{U}(n, \lceil n/2 \rceil)$ then with high probability, it is easy to see that $\Delta \notin \mathcal{M}(n, \lceil n/2 \rceil)$. Hence, by the definition of $\mathcal{M}(n, \lfloor n/2 \rfloor)$, with high probability, Δ^c has no faces of size $\lfloor n/2 \rfloor + 2$ or higher. Hence Δ has a complete $(\lceil n/2 \rceil - 2)$ -skeleton.

The proof is similar if $\Delta \sim \mathcal{U}(n, \lfloor n/2 \rfloor)$. \square

We use a combination of Chebyshev's Inequality and Markov's Inequality to prove a similar result on the skeletons of the random pure model, starting with the case $t = cn$.

Proposition 5.16. *Let $\Delta \sim \mathcal{RP}(n, t, p)$, let k be a constant and let $t' = t - k$. Suppose $t = cn$ where $0 < c < 1$ is a constant, and let ϵ be some positive constant.*

(a) *Let $k > 1$ and let $C = \frac{\alpha(c)k!}{(1-c)^k}$. If $p = \frac{C+\epsilon}{n^{k-1}}$, then Δ has a complete t' -skeleton with high probability. If instead $p = \frac{C-\epsilon}{n^{k-1}}$, then Δ with high probability does not have a complete t' -skeleton.*

(b) *Let $k = 1$. If $p = 1 - \exp\left(\frac{-\alpha(c)}{1-c}\right) + \epsilon$ then Δ has a complete t' -skeleton with high probability. If instead $p = 1 - \exp\left(\frac{-\alpha(c)}{1-c}\right) - \epsilon$, then Δ with high probability does not have a complete t' -skeleton.*

Before proceeding to the proof, we note that having a complete skeleton is a monotone property. More precisely, if $p < p'$ and $\Delta \sim \mathcal{RP}(n, t, p)$ and $\Delta \sim \mathcal{RP}(n, t, p')$ then if Δ has a complete r -skeleton with high probability, then so does Δ' .

Proof. For $x \in [n]^{(t')}$, we write X_x for the indicator of the event x is not a face. Let X denote the number of sets of size t' that are not faces of the complex, i.e. $X = \sum_x X_x$.

Note that X_x and X_y are independent if no set of size t contains x and y , i.e. whenever $|x \cap y| < t' - k$. Thus each A_x is independent from all but at most n^{2k} of the A_y , using the crude bound that there are fewer than n^k subsets of x of size $t' - k$, and each is contained in fewer than n^k sets of size t' .

If X_x and X_y are not independent, we use the easy bound, $\text{Cov}(X_x, X_y) \leq \mathbb{P}(X_x = 1)$.

Using Lemma 5.10, we have that

$$\begin{aligned}\text{Var}(X) &\leq \mathbb{E}(X) + n^{2k} \binom{n}{t'} (1-p)^{\binom{n-t'}{t-t'}} \\ &= (n^{2k} + 1) \mathbb{E}(X).\end{aligned}$$

(a) By linearity of expectation and using Proposition 5.13,

$$\begin{aligned}\mathbb{E}(X) &= \binom{n}{cn-k} (1-p)^{\binom{n-t'}{t-t'}} \\ &= \exp \left(\alpha(c)n + O(\log n) - p(1+o(1)) \frac{(n-cn)^k}{k!} \right),\end{aligned}$$

using the approximation $(1-p) = e^{p(1+o(1))}$, valid if $p = o(1)$.

It is easy to see that this tends to zero if $p = \frac{C+\epsilon}{n^{k-1}}$, and so in this case, by Markov's Inequality, Δ has a complete t' skeleton with high probability.

If instead $p = \frac{C-\epsilon}{n^{k-1}}$, then $\mathbb{E}(X)$ grows exponentially fast. Thus $\text{Var}(X) = o(\mathbb{E}(X)^2)$ and so Chebyshev's Inequality implies that $X > 0$ with high probability, i.e. the t' -skeleton is incomplete with high probability.

(b) Similarly,

$$\begin{aligned}\mathbb{E}(X) &= \binom{n}{cn-1} (1-p)^{n-cn+1} \\ &= \exp(\alpha(c)n + O(\log n)) (1-p)^{n-cn+1}.\end{aligned}$$

If $p = 1 - \exp\left(\frac{-\alpha(c)}{1-c}\right) + \epsilon$, this tends to zero and hence by Markov's Inequality, $X = 0$ with high probability.

If $p = 1 - \exp\left(\frac{-\alpha(c)}{1-c}\right) - \epsilon$, then once more, $\mathbb{E}(X) \rightarrow \infty$ exponentially quickly and so $\text{Var}(X) = o(\mathbb{E}(X)^2)$. Thus the t' -skeleton is, with high probability, not complete.

□

A special case of this is that $\mathcal{RP}(n, \frac{n}{2}, \frac{1}{2})$ has a complete $(n/2 - 2)$ -skeleton with high probability, agreeing with our earlier claim that $\mathcal{RP}(n, \frac{n}{2}, \frac{1}{2})$ is a good heuristic for $\mathcal{U}(n)$, when n is even.

Proposition 5.17. *Let $\Delta \sim \mathcal{RP}(n, t, p)$, where t is a constant and $t' < t$. If $p = (1 + \omega(n))^{\frac{t'(t-t')! \log n}{n^{t-t'}}}$, where $\omega(n) \log n \rightarrow \infty$, then with high probability, the t' -skeleton is complete.*

Proof. As before, let X denote the number of t' -sets that are not faces of Δ , and let p be as in the theorem's hypothesis.

$$\begin{aligned} \mathbb{E}(X) &= \binom{n}{t'} (1-p)^{\binom{n-t'}{t-t'}} \\ &\leq \exp \left(t' \log n - p \binom{n-t'}{t-t'} \right) \\ &\rightarrow 0. \end{aligned}$$

This concludes the proof. □

The case $t' = t-1$ tells us that when $p \geq (1 + \omega(n))^{\frac{(t-1) \log n}{n}}$, where $\omega(n) \log n \rightarrow \infty$, the random pure model agrees with the Linial-Meshulam model with respect to high probability events.

5.4 Homology Group

In this section, we study the top homology in both the uniform model and the random pure model. We start by proving Theorem 5.2. As in the proof of Proposition 5.15, we show that the result holds for $\Delta \sim \mathcal{U}(n, n/2, A, B)$ for any admissible pair (A, B) . By Korshunov's Theorem, this implies it also holds if $\Delta \sim \mathcal{U}(n)$.

Proof of Theorem 5.2. Let $\Delta \sim \mathcal{U}(n, n/2, A, B)$. We will show that for all fixed admissible pairs, (A, B) , the homology group $H_{n/2-1}(\Delta, \mathbb{F})$ has dimension at least $(1+o(1))2^{(n-1)/2}/\sqrt{\pi n}$ with high probability. By Theorem 5.5, this is enough to prove the result.

We let G be the collection of $(n/2 + 1)$ -sets, whose subsets of size $n/2$ are all free. In other words, $G = \{x \in [n]^{\binom{n}{n/2+1}} : x \setminus i \in F \text{ for all } i \in x\}$. Clearly, $|G| = (1+o(1))\binom{n}{n/2+1} = (1+o(1))2^{n+1/2}/\sqrt{\pi n}$. By definition of F , none of the sets in G are faces of Δ .

We say $x \in G$ is a *hole* if all subsets of x are present in Δ and let X denote the number of holes. We will use Chebyshev's Inequality to find a lower bound on X and then use this to prove a lower bound on $\dim(H_{n/2-1})$.

Let X_x be the indicator random variable that is 1 if x is a hole and 0 otherwise. Easily, $X = \sum_{x \in G} X_x$. It is easy to see that $\mathbb{E}(X_x) = (\frac{1}{2})^{n/2+1}$ and that $\mathbb{E}(X) = \binom{n}{n/2+1} 2^{-n/2-1} = (1 + o(1)) 2^{(n-1)/2} / \sqrt{\pi n}$.

If $x, y \in G$ have intersection of size at most $n/2 - 1$, then X_x and X_y are independent, and hence $\text{Cov}(X_x, X_y) = 0$.

If instead $|x \cap y| = n/2$, then x and y share a face of size $\frac{n}{2}$ and so

$$\begin{aligned} \text{Cov}(X_x, X_y) &= \mathbb{P}(X_x = 1, X_y = 1) - \mathbb{P}(X_x = 1)\mathbb{P}(X_y = 1) \\ &= 2^{-n-1} - 2^{-n-2} \\ &= 2^{-n-2}. \end{aligned}$$

Thus by Lemma 5.10, $\text{Var}(X) \leq \mathbb{E}(X) + n^2 |G| 2^{-n-2} = o(\mathbb{E}(X)^2)$ and so by Chebyshev's Inequality, $X = (1 + o(1)) 2^{(n-1)/2} / \sqrt{\pi n}$.

Each hole x can be identified naturally with a formal linear combination of its faces, v_x , namely $v_x := \sum_{i \in x} x \setminus \{i\}$. The v_x are each elements of $H_{\frac{n}{2}-1}$, by the definition of a hole.

We now show that almost all of these v_x are linearly independent, in fact, we prove the stronger statement that almost all the holes do not share facets with any other hole. We let Y_x be the indicator random variable that is 1 if x is a hole and and x that shares a face with another hole. In other words, $Y_x = 1$ if there exists an x' such that $|x \cap x'| = n/2$ and both x and x' are holes. We let $Y = \sum_x Y_x$. Easily, $\mathbb{P}(Y_x = 1) \leq n^2 2^{-n-1}$ and so by linearity of expectation, $\mathbb{E}(Y) \leq n^2 \binom{n}{n/2+1} 2^{-n-1}$. By Markov's Inequality,

$$\begin{aligned} \mathbb{P}\left(Y \geq \frac{1}{n} \mathbb{E}(X)\right) &\leq \frac{n \mathbb{E}(Y)}{\mathbb{E}(X)} \\ &\leq \frac{n^3 \binom{n}{n/2+1} 2^{-n-1}}{\binom{n}{n/2+1} 2^{-n/2-1}} \\ &= n^3 2^{-\frac{n}{2}}. \end{aligned}$$

This shows that with high probability, $X - Y \geq (1 + o(1)) 2^{(n-1)/2} / \sqrt{\pi n}$. Since $\dim(H_{n/2-1}) \geq X - Y$, this concludes the proof. \square

The same approach may be used to analyse the top homology of $\mathcal{U}(n, t, A, B)$, when n is odd and $t \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$ and thus prove the following analogue of Theorem 5.2—we omit the proof here.

Theorem 5.18. *If $\Delta \sim \mathcal{U}(n)$, n is odd and \mathbb{F} is a field, then with high probability, either $H_{\lfloor n/2 \rfloor - 1}(\Delta, \mathbb{F})$ has dimension at least $(1 + o(1))2^{\lceil n/2 \rceil - 1/2}$ or $H_{\lceil n/2 \rceil - 1}(\Delta, \mathbb{F})$ has dimension at least $(1 + o(1))2^{\lfloor n/2 \rfloor - 1/2}$.*

We believe that Theorem 5.2 is essentially best possible. Indeed, we propose:

Conjecture 5.2. *Let \mathbb{F} be a field and let $\Delta \sim \mathcal{U}(n)$, for even n . Let $H_{n/2-1} := H_{n/2-1}(\Delta, \mathbb{F})$. Then with high probability, $\dim(H_{n/2-1}) = (1 + o(1))2^{(n-1)/2}/\sqrt{\pi n}$.*

Similar methods allow us to prove an analogous result for the random pure model. We also see that because $\alpha(\frac{1}{2}) = \log 2$, the following result agrees with our heuristic that $\mathcal{RP}(n, \frac{n}{2}, \frac{1}{2})$ behaves similarly to $\mathcal{U}(n)$ when n is even.

Theorem 5.19. *Let \mathbb{F} be a field and let $\Delta \sim \mathcal{RP}(n, t, p)$ and write $H_{t-1} := H_{t-1}(\Delta, \mathbb{F})$. We also let ϵ be any positive constant.*

(a) *Let $t = cn$, for some constant $0 < c < 1$. If $p \geq e^{-\alpha(c)/c+\epsilon}$, then with high probability H_{t-1} is non-trivial. If also $p \leq 1 - \epsilon$, then $\dim(H_{t-1}) \geq e^{\alpha(c)n+O(\log n)}p^{cn}$, with high probability.*

(b) *Let $t \geq 1$ be a constant. If $np \rightarrow \infty$, then with high probability, H_{t-1} is non-trivial. If also $p \leq 1 - \epsilon$, then $\dim(H_{t-1}) \geq (1 + o(1))\frac{(np)^{t+1}}{(t+1)!}$, with high probability.*

Proof. For a set x of size $t + 1$, we let X_x be the indicator random variable that is 1 when x is a hole, that is all its subsets of size t are faces of Δ . We write $X = \sum X_x$. As in the proof of Theorem 5.2, we also let Y_x be the indicator random variable that is 1 if x is a hole and x shares a face with another hole, and we set $Y = \sum Y_x$. For both (a) and (b), we first bound X using Chebyshev's Inequality and then show $\dim(H_{t-1})$ follows the same bounds by bounding Y , the number of holes sharing a facet with another hole.

(a) Using Proposition 5.13,

$$\mathbb{E}(X) = \binom{n}{t+1} p^{t+1} = \exp\left(\alpha(c)n + O(\log n)\right) p^{cn}.$$

This tends to infinity exponentially fast if $p \geq e^{-\alpha(c)/c+\epsilon}$. It is easy to see that X_x and X_y are independent if $|x \cap y| \leq t-1$. On the other hand, if $|x \cap y| = t$, then $\text{Cov}(X_x, X_y) \leq p^{2t+1}$. Thus $\text{Var}(X) \leq \mathbb{E}(X) + n^2 \binom{n}{t+1} p^{2t+1} = o(\mathbb{E}(X)^2)$, by Proposition 5.10, and so by Chebyshev's Inequality, $X \geq \exp\left(\alpha(c)n + O(\log n)\right)p^{cn}$ with high probability.

It is easy to see that $\mathbb{P}(Y_x = 1) \leq n^2 p^{2t+1}$ and so by linearity of expectation, $\mathbb{E}(Y) \leq n^2 \binom{n}{t+1} p^{2t+1}$. By Markov's Inequality, $\mathbb{P}(Y \geq \frac{1}{n} \mathbb{E}(X)) \leq n^3 p^t$. If $p \leq 1 - \epsilon$, this tends to zero as $n \rightarrow \infty$. Since $\dim(H_{t-1}) \geq X - Y$, with high probability, $\dim(H_{t-1}) \geq \exp\left(\alpha(c)n + O(\log n)\right)p^{cn}$.

If $p \geq 1 - \epsilon$, we have that $\dim(H_{t-1}) \geq \exp\left(\alpha(c)n + O(\log n)\right)(1 - \epsilon)^{cn}$, since adding faces of size t cannot decrease $\dim(H_{t-1})$.

(b) In this case,

$$\mathbb{E}(X) = \binom{n}{t+1} p^{t+1} = (1 + o(1)) \frac{(np)^{t+1}}{(t+1)!},$$

which tends to infinity since np does.

Once again, X_x and X_y are independent if $|x \cap y| \leq t-1$, while if $|x \cap y| = t$, $\text{Cov}(X_x, X_y) \leq p^{2t+1}$. Thus, again, $\text{Var}(X) = o(\mathbb{E}(X)^2)$ and thus by Chebyshev's Inequality, $X = (1 + o(1)) \frac{(np)^{t+1}}{(t+1)!}$ with high probability. Once more, we let Y be the number of $(t+1)$ -sets x that are holes but share no facet with another hole. Again using Markov's Inequality, we may bound Y by $\frac{1}{n} \mathbb{E}(X)$ with high probability, as long as $p \leq 1 - \epsilon$. The remainder of the proof is identical to part (a). \square

Once again, we believe that these bounds on p are essentially tight.

Conjecture 5.3. *Let \mathbb{F} be a field and let $\Delta \sim \mathcal{RP}(n, t, p)$. We also let ϵ be any positive constant.*

(a) *Let $t = cn$, for some constant $0 < c < 1$. If $p \leq e^{-\alpha(c)/c-\epsilon}$, then with high probability, H_{t-1} is trivial.*

(b) *Let t be a constant. If $np \rightarrow 0$, then with high probability, H_{t-1} is trivial.*

5.5 Euler Characteristic

In this section, we prove Theorem 5.3, give high probability bounds on the Euler characteristic of $\mathcal{U}(n)$, again making heavy use of Theorem 5.5 in our approach. We then show that the Euler characteristic of $\mathcal{U}(n, n/2)$ is not too tightly concentrated, which goes some way towards providing a counterpart to Theorem 5.3.

Proof of Theorem 5.3. Let $\Delta \sim \mathcal{U}(n)$. Recall that, $\chi = \sum_{i=1}^n (-1)^{i-1} f_i$, where f_i denotes the number of faces of size i . By Proposition 5.14, we may assume that Δ has a complete $(n/2 - 2)$ -skeleton, and thus $f_i = \binom{n}{i}$, for $i \leq n/2 - 2$. By Theorem 5.5, we may assume that $\Delta \in \mathcal{M}(n, n/2, A, B)$, for some admissible A and B . Thus we also have that $f_i = 0$ for $i \geq n/2 + 2$ and that $f_{n/2+1} = |B| \leq 2^{n/2}$. It remains only to examine $f_{n/2}$ and $f_{n/2-1}$.

Let F denote the collection of free sets. Recall that since $|A|, |B| \leq 2^{n/2}$, we have that $\binom{n}{n/2} - (n/2 + 1)2^{n/2+1} \leq |F| \leq \binom{n}{n/2}$. Let X denote the number of free sets that are faces of Δ . It is a binomial random variable with parameters $|F|$ and $\frac{1}{2}$. It thus has mean $\frac{1}{2}|F|$ and so by Hoeffding's inequality,

$$\mathbb{P}\left(\left|X - \frac{1}{2}|F|\right| \geq 2^{n/2}\right) \leq \exp\left(-\frac{2^{n+1}}{|F|}\right) \rightarrow 0,$$

since $|F| \leq \binom{n}{n/2} = O\left(\frac{2^n}{\sqrt{n}}\right)$.

Combined with our bounds on $|F|$, it is easy to see that $|f_{n/2} - \frac{1}{2}\binom{n}{n/2}| \leq n2^{n/2+1}$ with high probability.

Let Y denote the number of $(n/2 - 1)$ -sets that are not faces of Δ . By considering Δ^c , as in the proof of Proposition 5.14, we may assume that $Y \leq 2^{n/2}$ and thus that $f_{n/2-1} \geq \binom{n}{n/2-1} - 2^{n/2}$.

Thus, with high probability,

$$\chi = \sum_{i=1}^{n/2-1} (-1)^{i-1} \binom{n}{i} + (-1)^{n/2-1} \frac{1}{2} \binom{n}{n/2} + O(n2^{n/2}).$$

Note that $\sum_{i=1}^{n/2-1} (-1)^{i-1} \binom{n}{i} + \frac{1}{2} (-1)^{n/2-1} \binom{n}{n/2} = \frac{1}{2} \sum_{i=0}^n (-1)^{i-1} \binom{n}{i} - 2$, by symmetry of $\binom{n}{i}$ around $i = n/2$. As there are equally many odd sized subsets of $[n]$ as even sized subsets, this sum evaluates to -2 , concluding the proof.

□

The odd case is only slightly more complicated. Suppose that $\Delta \sim \mathcal{U}(n, t)$ for $t \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$. Then it is easy to use the same method as above to prove that with high probability,

$$\chi(\Delta) = \sum_{i=1}^{t-1} (-1)^{i-1} \binom{n}{i} + \frac{1}{2} (-1)^{t-1} \binom{n}{t} + O(n2^{n/2}).$$

Using the identity $\sum_{i=1}^{j-1} (-1)^{i-1} \binom{n}{i} = (-1)^{j-1} \binom{n}{j} + 1$, which may be verified by induction on j , we get in both cases that with high probability,

$$\chi(\Delta) = (-1)^{\lceil n/2 \rceil} \frac{1}{2n} \binom{n}{\lceil n/2 \rceil} + O(n2^{n/2}).$$

We now prove an anti correlation result for $\mathcal{U}(n, n/2)$, the uniform distribution on the collection of ‘well-behaved’ complexes.

Theorem 5.20. *Let $\Delta \sim \mathcal{U}(n, n/2)$, for even n . For any natural k , $\mathbb{P}(\chi(\Delta) = k) \leq cn^{1/4} 2^{-n/2}$, for some constant c .*

Proof. We fix admissible (A, B) and let $\Delta \sim \mathcal{U}(n, n/2, A, B)$.

We again generate Δ by choosing free sets to be in Δ with probability $\frac{1}{2}$. However, we do this in three stages here. We will do this in such a way that the sets in the third stage do not add any $n/2 - 1$ sized faces to Δ .

In the first stage, we choose $F_1 \subseteq F$, by placing free sets in it with probability p , to be chosen later. We then choose each of these sets to be in Δ with probability $\frac{1}{2}$. At this point, almost all of the $(n/2 - 1)$ -sets are in Δ .

We define $F_3 \subseteq F \setminus F_1$ to be the collection of free sets whose subsets of size $n/2 - 1$ are all already faces of Δ . If x is a free set not in F_1 , then $\mathbb{P}(x \in F_3) = \left(1 - \left(1 - \frac{p}{2}\right)^{\frac{n}{2}}\right)^{\frac{n}{2}}$. Thus:

$$\mathbb{E}(|F_3|) = (1 - p)|F| \left(1 - \left(1 - \frac{p}{2}\right)^{n/2}\right)^{n/2}.$$

Choose $p = \frac{3 \log n}{n}$. Then $\mathbb{E}(|F_3|) = (1 + o(1)) \binom{n}{n/2}$, and hence $\mathbb{E}(|F_2|) = o\left(\binom{n}{n/2}\right)$. It is easy to show also that $\text{Var}(|F_3|) = o(\mathbb{E}(|F_3|))$. Thus, by Chebyshev’s Inequality, $\mathbb{P}\left(|F_3| < \frac{1}{2} \binom{n}{n/2}\right) \leq \binom{n}{n/2}^{-1}$. Recall that $\binom{n}{n/2} = \Theta\left(\frac{2^n}{\sqrt{n}}\right)$, and so we may assume this does not happen.

Let $F_2 = F \setminus (F_1 \cup F_3)$. We choose sets in F_2 to be faces of Δ with probability $\frac{1}{2}$, then complete the construction of Δ by repeating this process for sets in F_3 .

We see that the probability of the Euler characteristic being a given value is at most the probability of $|F_3|$ fair coins having a given number of heads. This is at most $1/\sqrt{|F_3|} = \frac{c \cdot n^{1/4}}{2^{n/2}}$, and so we may conclude. □

The same methods also show that if n is odd and $\Delta \sim \mathcal{U}(n, t)$, for $t \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$, the conclusion of the theorem still holds, i.e. for any natural k , $\mathbb{P}(\chi(\Delta) = k) \leq cn^{1/4}2^{-n/2}$, for some constant c .

A very easy corollary of Theorem 5.20 and Theorem 5.5 is the following, which we will use in Section 5.7.

Corollary 5.21. *Let $k = k(n)$ be any integer function of n and let $\Delta \sim \mathcal{U}(n)$. With high probability, $\chi(\Delta) \neq k$.*

5.6 Shellability and the h -Vector

We say a complex is *shellable* if its facets can be arranged in an order, x_1, \dots, x_m in such a way that for all $i < j$, if $x_i \cap x_j$ is non-empty, then there is some $k \leq j$ for which $x_i \cap x_j \subseteq x_k \cap x_j$ and $|x_k \cap x_j| = |x_j| - 1$. Such an order is called a *shelling*. See [10] for a broader introduction. Among other things, they show that if Δ is shellable, then it has a shelling with facets in non-increasing order of size.

Although interesting in its own right, part of the reason for studying shellable complexes is that they are also Cohen-Macaulay complexes. See [87] or [20] for background on these well-studied complexes.

Theorem 5.22. *For all even n , if $\Delta \sim \mathcal{U}(n)$ then with high probability, Δ is not shellable.*

Proof. By Theorem 5.5 it is sufficient to prove this for $\Delta \sim \mathcal{U}(n, n/2)$. As mentioned above, if Δ is shellable, then there is a shelling with facets in non-increasing order of size. Thus if Δ has at least two bundles in the $(i+1)^{th}$ layer, Δ is not shellable. By Theorem 5.6, this happens with high probability. □

The situation is significantly more complicated for the random pure model, and in particular, we are not quite able to show $\mathcal{RP}(n, \frac{n}{2}, \frac{1}{2})$ is not shellable, although we show $\mathcal{RP}(n, \frac{n}{2}, \frac{1}{2} - \epsilon)$ is not shellable with high probability, for all positive constants ϵ .

Theorem 5.23. *Let $\Delta \sim \mathcal{RP}(n, t, p)$, and $\epsilon > 0$ be any constant.*

(a) *Let $t = cn$, for some constant c . If $n^{-2+\epsilon} < p < 1 - \exp\left(-\frac{\alpha(c)}{2(1-c)}\right) + \epsilon$, then with high probability, Δ is not shellable.*

(b) *Let $t > 2$ be a constant. If $n^{-t+\epsilon} < p < (\frac{t-1}{2} - \epsilon)\frac{\log n}{n}$ then with high probability, the simplex is not shellable.*

Proof. Given Δ and a $(t-2)$ -set, x , we write G_x for the graph whose vertex set is the set of facets of Δ that contain x , and an edge is present between two facets if their intersection is a set of size $t-1$. We call this the *x -intersection graph* of Δ .

Note that if G_x is not connected, then Δ is not shellable (a graph with no vertices is here considered to be connected). This is because in a shelling, after the first facet that contains x , each subsequent facet that contains x must have an intersection of size $t-1$ with a previously added facet that contains x . In particular, if G_x has an isolated vertex, Δ is not shellable, unless G_x is a single vertex.

Let $N_{x,v}$ be the indicator random variable that is 1 if and only if the vertex v in the graph G_x is isolated.

We write N_x for the number of isolated vertices in the graph G_x , and N for the total number of isolated vertices in all of these graphs. That is, $N_x = \sum_v N_{x,v}$ and $N = \sum_x N_x$. It is easy to see that:

$$\mathbb{E}(N_x) = \binom{n-t+2}{2} p(1-p)^{2(n-t)}.$$

Since $N_{x,v}$ depends only on the presence of t -sets u for which $|u \cap v| \geq t-1$, the random variables $N_{x,v}$ and $N_{x',v'}$ are independent unless $|v \cap v'| \geq t-2$.

(a) Using Proposition 5.13, we see that:

$$\begin{aligned} \mathbb{E}(N) &= \binom{n}{cn-2} \binom{n-cn+2}{2} p(1-p)^{2(1-c)n} \\ &= \exp(\alpha(c)n + O(\log n)) p(1-p)^{2(1-c)n}. \end{aligned}$$

Note that this is unimodal as a function of p , and hence its lowest value in the range of p in the theorem is attained by one of the endpoints of the range. Thus for p in the range of the theorem, this tends to infinity exponentially fast. Using the bound $\text{Cov}(N_{x,v}, N_{x',v'}) \leq \mathbb{P}(N_{x,v} = 1) = p(1-p)^{2(n-t)}$

$$\begin{aligned}\text{Var}(N) &\leq \mathbb{E}(N) + n^4 \binom{n-cn+2}{2} \binom{n}{cn-2} p(1-p)^{2(1-c)n} \\ &= (1+n^4)\mathbb{E}(N) \\ &= o(\mathbb{E}(N)^2).\end{aligned}$$

Thus Chebyshev's Inequality implies that $N = (1 + o(1))\mathbb{E}(N)$ with high probability.

Let M_x denote the indicator random variable that is 1 if G_x consists of only one vertex, and let $M = \sum_x M_x$.

$$\mathbb{E}(M) = \binom{n}{cn-2} \binom{n-cn+2}{2} p(1-p)^{\binom{n-cn+2}{2}-1}.$$

Since $p > n^{-2+\epsilon}$, this is a smaller order of magnitude than $\mathbb{E}(N)$, so by Markov's Inequality, $M = o(N)$ with high probability. Thus with high probability, $N - M > 0$ and so the complex is not shellable.

(b) We again use the linearity of expectation to see that:

$$\begin{aligned}\mathbb{E}(N) &= (1 + o(1)) \frac{n^{t-2}}{(t-2)!} \frac{n^2}{2} p(1-p)^{2(n-t)} \\ &= (1 + o(1)) \frac{n^t}{2(t-2)!} p \exp(-2p).\end{aligned}$$

For p in the range of the theorem, this tends to infinity. This time we require a slightly more delicate argument to bound the covariances. We make use of the fact that $\text{Cov}(N_{x,v}, N_{x',v'}) \leq \mathbb{P}(N_{x,v} = 1 \text{ and } N_{x',v'} = 1) \leq (1-p)^{4(n-t)-f(x,x',v,v')}$, where $f(x, x', v, v')$ denotes the number of t -sets, w , such that $x, x' \subseteq w$, $|v \cap w| = t-1$ and $|v' \cap w| = t-1$. If $|v \cap v'| = t-2$, it is easy to see that $f(x, x', v, v') \leq 4$. If $|v \cap v'| = t-1$, we can see that $f(x, x', v, v') \leq n$; again, we do not require the condition that $x, x' \subseteq w$ for this bound. Suppose instead that $v = v'$. We assume $x \neq x'$ (otherwise the covariance is zero). Then

$|x \cup x'| \geq t - 1$, so we have at most $n - t$ choices for w , and so $f(x, x', v, v') \leq n - t$. Hence,

$$\begin{aligned} \text{Var}(N) &\leq \mathbb{E}(N) + t^2 n^2 \binom{n-t+2}{2} \binom{n}{t-2} p^2 (1-p)^{4n-4t-4} \\ &\quad + tn \binom{n-t+2}{2} \binom{n}{t-2} p^2 (1-p)^{3n-4t} \\ &= o(\mathbb{E}(N)^2). \end{aligned}$$

Thus by Chebyshev's Inequality, $N = (1 + o(1))\mathbb{E}(N)$ with high probability.

Let M_x denote the indicator random variable that is 1 if G_x consists of only one vertex, and let $M = \sum_x M_x$.

$$\mathbb{E}(M_x) = \binom{n}{t-2} \binom{n-t+2}{2} p(1-p)^{\binom{n-t+2}{2}-1}.$$

Once again, this is a smaller order of magnitude than $\mathbb{E}(N)$, so by Markov's Inequality, $M = o(N)$ with high probability. Thus with high probability, $N - M > 0$ and so the complex is not shellable. \square

We now introduce an important concept related to Cohen-Macaulay complexes. For a pure simplex Δ , with facets of size t , the h -vector $h(\Delta) = (h_0(\Delta), h_1(\Delta), \dots, h_t(\Delta))$ is defined by the relation $h_k = \sum_{i=0}^k (-1)^{k-i} \binom{t-i}{k-i} f_i$, where, as before, f_i denotes the number of faces of size i . In particular, $h_t = (-1)^t \sum_{i=0}^t (-1)^i f_i$, and is a linear translate of the Euler characteristic.

In the range of large t , we are able to find ranges of p for which $h_t(\Delta) < 0$ with high probability, if $\Delta \sim \mathcal{RP}(n, t, p)$. We note that all Cohen-Macaulay complexes (CM complexes) have positive h -vector, so this gives a lower bound on the threshold for being a CM complex

Theorem 5.24. *Let $\Delta \sim \mathcal{RP}(n, t, p)$, let ϵ be a positive constant and let $t = cn$ for some constant $0 < c < 1$. If $\frac{2\alpha(c)+\epsilon}{2(1-c)^2n} < p < c - \epsilon$, then with high probability, $h_t < 0$.*

Proof. We will treat both cases together and then use the lower bounds on p to conclude using Propositions 5.16 and 5.17.

Firstly, we note that f_t is binomially distributed, with parameters $\binom{n}{t}$ and p . Thus $\mathbb{E}(f_t) = p \binom{n}{t}$ and $\text{Var}(f_t) = p(1-p) \binom{n}{t} = o(\mathbb{E}(f_t)^2)$, so Chebyshev's Inequality allows us to conclude that with high probability, $f_t = (1 + o(1))p \binom{n}{t}$.

For sets $x \in [n]^{(t-1)}$, we write X_x for the indicator random variable that is 1 if x is not a face of Δ and let $X = \sum_x X_x$. Clearly, $\mathbb{E}(X) = \binom{n}{t-1}(1-p)^{n-t+1}$. Since $f_{t-1} = \binom{n}{t-1} - X$, we may use Markov's Inequality to show that with high probability, $f_{t-1} = (1+o(1))\binom{n}{t-1}$.

By Propositions 5.16 and 5.17, with high probability, $f_i = \binom{n}{i}$ for all $i \leq t-2$. Thus with high probability,

$$h_t = \sum_{i=0}^{t-1} (-1)^{t-i} \binom{n+1}{i} + (1+o(1))p \binom{n}{t} + o\left(\binom{n}{t-1}\right).$$

Since $\sum_{i=0}^j (-1)^{j-i} \binom{n}{i} = \frac{j+1}{n} \binom{n}{j+1}$, as seen by induction on j , the above simplifies to $h_t = ((1+o(1))p - \frac{t}{n}) \binom{n}{t} + o\left(\binom{n}{t-1}\right)$, which concludes the proof. \square

We note that since h_t is positive if Δ is shellable, this theorem extends the range for non-shellability given by Theorem 5.23. Indeed, we have:

Corollary 5.25. *Let $\Delta \sim \mathcal{RP}(n, t, p)$, and $\epsilon > 0$ be any constant. Let $t = cn$, for some constant c . If $n^{-2+\epsilon} < p < 1 - \exp\left(-\frac{\alpha(c)}{2(1-c)}\right) + \epsilon$, then with high probability, Δ is not shellable.*

5.7 Evasiveness

In this section, we prove Corollary 5.4, using a previously established connection between evasive functions and a topological property of simplicial complexes known as collapsibility.

A face x of Δ is known as a *removable face* if it is not a facet and there is only one facet containing it. A *collapse* of Δ is a simplicial complex formed by deleting from Δ all the faces containing some fixed removable face, x , including x itself. For example, if $\Sigma = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}\}$, then $\{3\}$ is a removable face and $\Sigma' = \{\{1\}, \{2\}, \{1,2\}\}$ is a collapse of Δ . We say Δ is *collapsible* if there is a sequence of collapses that reduces Δ to a single size one face. More precisely, Δ is collapsible if there is a sequence $\Delta_1, \dots, \Delta_t$ such that Δ_{i+1} is a collapse of Δ_i , for $1 \leq i \leq t-1$, and such that $\Delta_1 = \Delta$ and $|\Delta_t| = 1$. It

is easy to see that the previous example, Σ , is collapsible. Note that each collapse preserves the Euler characteristic, so all collapsible complexes have Euler Characteristic 1.

For $A \subseteq [n]$, let x_A denote the Boolean string whose i^{th} co-ordinate is 1 if and only if $i \in A$. We write Δ_f for the subset of $\mathcal{P}[n]$ that consists of all non-empty subsets A such that $f(x_A) = 0$. If f is a monotone function, Δ_f is a simplicial complex. Note that every simplicial complex may be associated with a monotone Boolean function in this way. The usefulness of this viewpoint when studying evasiveness is demonstrated by the following lemma.

Lemma 5.26 (Kahn-Saks-Sturtevant [55]). *If f is a non-constant monotone Boolean function that is not evasive, then Δ_f is collapsible.*

Proof of Corollary 5.4. It is sufficient to show that almost all simplicial complexes are not collapsible. Since collapsible complexes have trivial homology group, this follows from Theorems 5.2 and Proposition 5.18. \square

We note that that the result may alternatively be derived from Corollary 5.21 on the Euler characteristic of the uniform complex.

Bibliography

- [1] N. Alon *Neighborly families of boxes and bipartite coverings* Algorithms and Combinatorics **14** (1997) 27–31.
- [2] N. Alon, A. Krech, T. Szabò. *Turán’s theorem in the hypercube*. SIAM J. Discrete Maths. **21**(1) (2007) 66–72.
- [3] N. Alon and J. H. Spencer, *The Probabilistic Method*, John Wiley and Sons, New York, (2008). ISBN: 978-0470170205.
- [4] L. Aronshtam, N. Linial, T. Łuczak and R. Meshulam, *Collapsibility and Vanishing of Top Homology in Random Simplicial Complexes*, Discrete Comput. Geom., **49**, (2013) 317–334.
- [5] L. Aronshtam and N. Linial, *When does the top homology of a random simplicial complex vanish?* Random Structures Algorithms, **46**(1), (2015), 26–35.
- [6] E. Babson, C. Hoffman and M. Kahle, *The fundamental group of random 2-complexes*, *J. Amer. Math. Soc.*, **24**(1) (2011), 1–28.
- [7] J. Balogh, B. Bollobás, R. Morris and O. Riordan, *Linear algebra and bootstrap percolation*, J. Combin. Theory Ser. A, **119**(6) (2012), 1328–1335.
- [8] J. Balogh, P. Hu, B. Lidický and H. Liu. *Upper bounds on the size of 4- and 6-cycle-free subgraphs of the hypercube*. European J. Combin. **35** (2014) 75–85.
- [9] A. J. Bernstein, *Maximally connected arrays on the n -cube*, SIAM J. Appl. Math., **15** (1967) 1485–1489.

- [10] A. Björner and M. Wachs, Shellable Nonpure Complexes and Posets. I, *Trans. Amer. Math. Soc.*, **348**(4), (1996).
- [11] T. Bohman, *The Triangle-Free Process*, Advances in Mathematics, **221** (2009) 1653-1677.
- [12] T. Bohman, P. Keevash, *The early evolution of the H -free process*, *Inventiones Mathematicae*, **181** (2010), 291–336.
- [13] T. Bohman, P. Keevash, *Dynamic concentration of the triangle-free process*, arxiv: 1302.5963, (2013).
- [14] B. Bollobás, *On a conjecture of Erdős, Hajnal and Moon*, *Amer. Math. Monthly*, **71**, (1984) 178–179.
- [15] Béla Bollobás, *Random Graphs*, Academic Press, Inc., London, **1985**, ISBN: 0-12-111755-3; 0-12-111756-1
- [16] B. Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics, **184**, Springer-Verlag, New York, 1998, ISBN: 0-387-98488-7.
- [17] B. Bollobás and I. Leader, *Matchings and paths in the cube*, *Discrete Appl. Math.*, **75** (1997) 1–8.
- [18] P. Brass, H. Harborth and H. Nienborg. *On the maximum number of edges in a C_4 -free subgraph of Q_n* . *J. Graph Theory*. **19** (1) (1995) 17–23.
- [19] A. Brooke-Taylor and D. Testa, , *The infinite random simplicial complex*, ArXiv e-prints 1308.5517.
- [20] W. Bruns and J. Herzog, *Cohen-Macaulay Rings*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, **39**, (1993).
- [21] N. G. De Bruijn and P. Erdős, *On a combinatorial problem*, *Nederl. Akad. Wetensch. Proc.*, **51** (1948), 1277–1279
- [22] P. J. Cameron, *Problems from the 16th British Combinatorial Conference*, *Discrete Math.*, **293** (2005) no. 1–3, 313–320.

- [23] S. Choi and P. Guan. *Minimum Critical Squarefree Subgraph of a Hypercube*. Proceedings of the Thirty-Ninth Southeastern International Conference on Combinatorics, Graph Theory and Computing. **189** (2008) 57–64.
- [24] F.R.K. Chung *On the coverings of graphs* Discrete Mathematics **30** 89–93.
- [25] F. R. K. Chung, P. Erdős and J. Spencer *On the decomposition of graphs into complete bipartite subgraphs* Studies in Pure Mathematics, Mem. of P. Turán (1983) 95–101.
- [26] F. R. K. Chung, Z. Füredi, R. L. Graham and P. D. Seymour, *On induced subgraphs of the cube*, J. Combin. Theory Ser. A, **49** (1988) 180–187.
- [27] L. Csirmaz, P. Ligeti and G. Tardos, *Erdős–Pyber Theorem for Hypergraphs and Secret Sharing*, Graphs and Combinatorics, **31**(5) (2015) 1335–1346.
- [28] R. Diestel, *Graph Theory*, Graduate Texts in Mathematics, Volume 173 Springer-Verlag, Heidelberg, 2005 ISBN 978-3-642-14278-9.
- [29] J. Dong and Y. Liu, *On the decomposition of graphs into complete bipartite graphs*, Graphs and Combinatorics, **23** (2007) no. 3, 255–262.
- [30] P. Erdős *Graph theory and probability* Canadian Journal of Mathematics **11** (1959) 34–38.
- [31] P. Erdős. *Some problems in graph theory, combinatorial analysis and combinatorial number theory*. In Graph Theory and Combinatorics. B. Bollobás, ed. Academic Press, London. (1984) 1–17.
- [32] P. Erdős, A. Goodman and L. Pósa. *The representation of graphs by set intersections*, Canad. J. Math., **18** (1966), 106–112
- [33] P. Erdős, A. Hajnal and J. W. Moon. *A Problem in Graph Theory*. The American Mathematical Monthly. **71**(10) (1964) 1107–1110.
- [34] P. Erdős and L. Pyber, *Covering a graph by complete bipartite subgraphs*, Discrete Mathematics, **170** (1997) 249–251.

- [35] J. R. Faudree, R. J. Faudree and R. Schmitt. *A Survey of Minimum Saturated Graphs*, The Electronic Journal of Combinatorics. **18** (2011).
- [36] P.C. Fishburn and P.L. Hammer *Bipartite dimensions and bipartite degrees of graphs*, Discrete Mathematics, **160** (1996) Issues 1–3, 127–148.
- [37] L. R. Ford and D. R. Fulkerson, *Maximal flow through a network*, Canad. J. Math., **8** (1956), 399–404.
- [38] K. Gandhi, C. Kim, *Saturation Number of Trees in the Hypercube*, arxiv: 1409.7983, (2014).
- [39] R. L. Graham and H.O. Pollak, *On the addressing problem for loop switching*, Bell System Technical Journal, **50** (1971) 2495–2519.
- [40] A. Gundert and U. Wagner, *On topological minors in random simplicial complexes*, Proc. Amer. Math. Soc., **144** (4), (2016), 1815–1828.
- [41] P. Hall, *On Representatives of Sets*, J. London Math. Soc., **10** (1), (1935), 26–30.
- [42] R. Hamming, *Error detecting and error correcting codes*, Bell System Technical Journal, **29**(2) (1950), 147–160.
- [43] F. Harary, *Graph Theory*, Addison–Wesley, Reading, Massachusetts, 1969, ISBN: 0-201-41033-8.
- [44] L. H. Harper, *Optimal assignments of numbers to vertices*, SIAM J. Appl. Math., **12** (1964) 131–135.
- [45] L. H. Harper, *Optimal numberings and isoperimetric problems on graphs*, J. Combin. Theory, **1** (1966) 385–394.
- [46] S. Hart, *A note on the edges of the n -cube*, Discrete Math., **14** (1976) 157–163.
- [47] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, **58**(301), 13–30.
- [48] J.R. Johnson and K. Markström, *Turán and Ramsey Properties of Subcube Intersection Graphs*, Combinatorics, Probability and Computing **22** (2013) 55–70.

- [49] J. R. Johnson and T. Pinto, *Saturated Subgraphs of the Hypercube*, arxiv 1406.1766, (2014).
- [50] J. R. Johnson and J. Talbot, *Vertex Turán problems in the hypercube*. Journal of Combinatorics (Series A). **117** (2010), 454–465.
- [51] S. Jukna, *On Set Intersection Representations of Graphs* Journal of Graph Theory **61** (1) (2009) 55–75.
- [52] S. Jukna, *Boolean Function Complexity*, Algorithms and Combinatorics, **27**, Springer–Verlag, ISBN: 978-3-642-24507-7.
- [53] S. Jukna and A. Kulikov, *On covering graphs by complete bipartite subgraphs*, Discrete Mathematics, **309** (10) (2009) 3399–3403.
- [54] S. Jukna, I. Sergeev, *Complexity of Linear Boolean Operators*, Foundations and Trends in Theoretical Computer Science, **9**(1) (2013) 1–123.
- [55] J. Kahn, M. Saks and D. Sturtevant, A topological approach to evasiveness, *Combinatorica*, **4**(4) (1984), 297–306.
- [56] L. Kászonyi and Z. Tuza, *Saturated graphs with minimal number of edges*, J. Graph Theory, **10**(2) (1986) 203–210.
- [57] G. Katona and E. Szemerédi, *On a problem of graph theory*, Studia Scientiarum Mathematicarum Hungarica, **2** (1967) 23–28.
- [58] G. O. H. Katona and T. G. Tarján. *Extremal problems with excluded subgraphs in the n -cube*. Graph Theory, Lagow, Poland, Lecture Notes in Mathematics. **1018**, Berlin: Springer (1983), 84–93.
- [59] A. D. Korshunov, On the number of monotone Boolean functions, *Problemy Kibernetiki*, **38**, (1981) 5–108 (in Russian).
- [60] A. D. Korshunov, Monotone Boolean functions, Russian Math. Surveys, **58**(5) (2003), 929–1001.

- [61] A.D. Korshunov and I. Shmulevich, On the distribution of the number of monotone Boolean functions relative to the number of lower units, *Discrete Mathematics*, **257**, (2002) 463–579.
- [62] D. Korándi and B. Sudakov, *Saturation in random graphs*, arXiv: 1510.09187 (2015).
- [63] E. Kostochka, *Piercing the edges of the n -dimensional unit cube*, Diskret. Analiz Vyp., **28** (1976), 55–64 (in Russian).
- [64] R. Kulkarni, Y. Qiao and X. Sun, *Any Monotone Property of 3-uniform Hypergraphs is Weakly Evasive*, Electronic Colloquium of Computational Complexity, **61** (2014).
- [65] E. Kushilevitz and N. Nisan *Communication Complexity* Cambridge University Press (1997).
- [66] N. Linial, R. Meshulam, *Homological Connectivity of Random 2-Complexes*, Combinatorica, **26** (4) (2006) 475–487.
- [67] J. H. Lindsey, *Assignment of numbers to vertices*, Amer. Math. Monthly, **71** (1964) 508–516.
- [68] L. Lovász, *Combinatorial Problems and Exercises*, pg81 (ex 13.31), North-Holland Publishing Co., Amsterdam, 1993. ISBN: 0-444-81504-X 05-01.
- [69] O.B. Lupanov *On rectifier and switching-and-rectifier schemes*, Doklady Akademii Nauk SSSR **111** (1956) 1171–1174 (in Russian).
- [70] F. H. Lutz, Some results related to the evasiveness conjecture, J. Combin. Theory Ser. B **81** (2001), no. 1, 110–124.
- [71] T.A. McKee and F.R. McMorris *Topics in Intersection Graph Theory* SIAM Monographs on Discrete Mathematics and Applications (1999).
- [72] K. Menger, *Zur allgemeinen Kurventheorie*, Fund. Math., **10** (1927), 96–115 (in German).
- [73] R. Meshulam, N. Wallach, *Homological Connectivity of Random k -dimensional Complexes*, Random Structures & Algorithms, **34**(3) (2009), 408–417, May 2009.

- [74] M. Mehta, *Random Matrices*, Pure and Applied Mathematics, Elsevier/Academic Press, Amsterdam **142** (2004), ISBN 0-12-088409-7.
- [75] N. Morrison, J. A. Noel and A. Scott, *On Saturated k -Sperner Systems*, arXiv: 1402.5646, (2014).
- [76] N. Morrison, J. A. Noel and A. Scott, *Saturation in the Hypercube and Bootstrap Percolation*, arXiv: 1408.5488, (2014).
- [77] D. Mubayi and S. Vishwanathan *Bipartite Coverings and the Chromatic Number* Electronic Journal of Combinatorics **16** (2009) N34.
- [78] D.S. Nau, G. Markowski, M. A. Woodbury and D. B. Amos *A Mathematical Analysis of Human Leukocyte Antigen Serology* Mathematical Biosciences **40** (1978) 243–270.
- [79] A. Nieves Roman, *Subgrafos Criticos Sin Cuadrados En Hipercubos*, Master Thesis, Universidad de Puerto Rico (2000) (in Spanish).
- [80] O. Pikhurko, *Results and Open Problems on Minimum Saturated Hypergraphs*. Ars Combinatoria. **72** (2004) 435–451.
- [81] T. Pinto, *Biclique covers and partitions*, Electron. J. Combin., **21**(1) (2014).
- [82] T. Pinto, *The proofs of two directed paths conjectures of Bollobás and Leader*, arXiv:1504.07079.
- [83] G. Pontiverios, S. Griffiths and R. Morris, *The triangle-free process and $R(3, k)$* , arXiv: 1302.6279, (2013).
- [84] R. Rivest, J. Vuillemin, *On Recognizing Graph Properties from Adjacency Matrices*, *Theoretical Computer Science*, **3** (1976) 371–384.
- [85] V. Rödl and A. Ruciński, *Bipartite Coverings of Graphs*, *Combinatorics, Probability and Computing* , **6** (1997) 349–352.
- [86] A. Ruciński and N. Wormald, *Random graph processes with degree restrictions*, *Combinatorics, Probability and Computing*, **1** (1992) 169–180.

- [87] R. P. Stanley, *Balanced Cohen-Macaulay Complexes*, Trans. Amer. Math. Soc., **249** (1), (1979), 139–157.
- [88] H. Tuza, *Covering of graphs by complete bipartite subgraphs; Complexity of 0-1 matrices*, Combinatorica, **4** (1984) 111–116.
- [89] H. Tverberg, *On the Decomposition of K_n into Complete Bipartite Graphs*, Journal of Graph Theory, **6** (1982) 493–494.
- [90] J. van Lint, *Introduction to Coding Theory*. Berlin; Springer-Verlag (1999).
- [91] W. Wessel, *Über eine Klasse paarer Graphen, I: Beweis einer Vermutung von Erdős, Hajnal und Moon*, Wiss. Z. Hochsch. Ilmenau, **12** (1966), 253-256 (in German).
- [92] N. Wormald, *The differential equations method for random graph processes and greedy algorithms*. Lectures on Approximation and Randomized Algorithms, 73–155, Karonski and Prömel eds. PWN, Warsaw 1999.
- [93] A. Zykov, *On some properties of linear complexes*, Mat. Sbornik. N. S., **24** (1949), 163–168 (in Russian).